Expansion of Texas Land Use/Land Cover through Class Crosswalking and Lidar Parameterization of Arboreal Vegetation

TCEQ Grant # 582-5-64593-FY09-25

Principle Investigator:
Sorin C. Popescu, Ph.D.
Associate Professor
Department of Ecosystem Science & Management
Texas A&M University

Secondary Investigators:
Jared Stukey
Muge Mutlu, Ph.D.
Kaiguang Zhao, Ph.D.
Ryan Sheridan
Nian-Wei Ku

TCEQ Project Manager:
Clint Harper
Air Quality Division

Abstract

The Expansion of Texas Land Use/Land Cover through Class Crosswalking and light detection and ranging (lidar) Parameterization of Arboreal Vegetation project was initiated by the Texas Commission on Environmental Quality to provide a more detailed and accurate map of land cover necessary for air quality modeling for the 12km Comprehensive Air Quality Model with Extensions (CAMx) domain. The project consisted of crosswalking classes from the LANDFIRE and Texas Parks and Wildlife Vegetation classes and classifying LandSat imagery to the Texas Land Classification System, and to derive forest composition characteristics with lidar for more accurate biogenic emission modeling. Lidar was used to estimate tree height, canopy base height, diameter at breast height, individual tree biomass, and canopy bulk density. Individual trees were identified through using lidar and the TreeVaw software, which uses a local maxima varying filter.

Table of Contents

Description of the Grant Activity

The grant activity description received from the TCEQ was written as an itemized list of 5 overall tasks and a total of 21 subtasks (Table1). Tasks 1 and 2 required the PI to return an approved grant activity description (work plan) and file monthly progress reports, respectively. Task 5 addresses the issuance of this Final Report. Discussion of these tasks is unnecessary and is not included in the report. This report is organized into seven chapters which detail the steps taken to complete the tasks. Some chapters encompassing several tasks and some address a single task. Task references, including a full description of tasks, are found in APPENDIX A.

Task	Description
Task 3 – Identification of LULC classes that have analogues with the NLCD vs. classes without direct correlation to NLCD:	Identification of LULC classes that have analogues with the NLCD vs. classes without direct correlation to NLCD
Task 3.1 – Identification and integration of ancillary GIS data:	Identification and integration of ancillary GIS data
Task 3.2	Field Work
Task 3.3 – Supplemental Imagery Classification:	Supplemental Imagery Classification
Task 3.4 – Conversion Routines for converting NLCD to TX LULC	Conversion Routines for converting NLCD to TX LULC
Task 3.5 – Final Classification and Shapefiles:	Final Classification and Shapefiles
Task 4.0 – Canopy Height Model	Canopy Height Model
Task 4.1 – Above-ground Biomass Map:	Above-ground Biomass Map
Task 4.2 – Surface Roughness:	Surface Roughness
Task 4.3 – Ground-based LiDAR field measurements:	Ground-based lidar field measurements
Task 4.4 – Maps of individual tree crown dimensions:	Maps of individual tree crown dimensions
Task 4.5 – Comparison of ground and airborne LiDAR measurements compared	Comparison of ground and airborne lidar measurements compared
4.6	Crown bulk density or LMD
Task 4.7 – Investigation of ground-based methodology for LAI and LMD assessment and comparison to airborne data:	Investigation of ground-based methodology for LAI and LMD assessment and comparison to airborne data
Task 4.8 – 2009 LiDAR Acquisition:	2009 Lidar Acquisition
Task 4.9 – Hyperion Acquisition:	Hyperion Acquisition

Table 1. Itemized list of Tasks and a brief description of each.

Introduction

The Texas Land Use/Land Cover (TX LULC) map was generated, with funding from the Texas Commission on Environmental Quality (TCEQ), in order to provide a more detailed and accurate map of land cover necessary for air quality modeling. At present, the TX LULC extends far enough to fulfill the TCEQ's modeling needs in the East Texas 4km Comprehensive Air Quality Model with Extensions (CAMx) domain. This project extends the TX LULC classes to the 12km CAMx domain (Figure 1) by "crosswalking" (mapping one classification to another) from similar LANDFIRE classes. Because the 12km domain covers a large area of the southeastern United States, this "crosswalking" is complicated by changes in vegetation composition within each class resulting from climatological heterogeneity. The project includes fieldwork and detailed image analysis necessary to reconcile changes in vegetation across the domain (i.e., the NLCD category "broadleaf forest" in Texas may not contain the same composition of trees as "broadleaf forest" in Florida). This crosswalk will enable future LULC input files to be quickly created from LANDFIRE data for the TCEQ biogenic emissions model, Global Biosphere Emissions and Interactions System (GloBEIS).

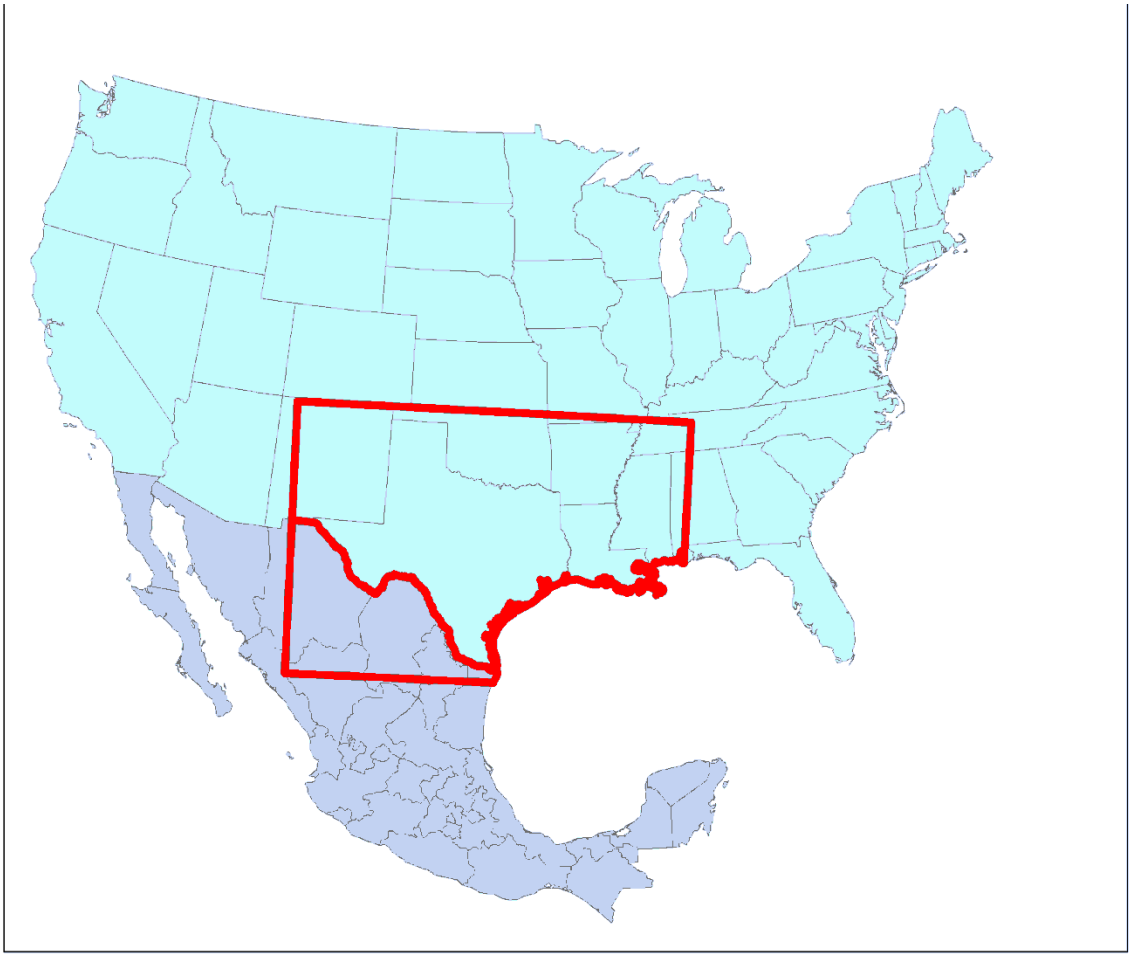


Figure 1. Extent of the 12km CAMx domain with the border between the United States and Mexico indicating the separation between LANDFIRE and LandSat data used for classification.

Biogenic emissions are affected by Leaf Mass Density (LMD), species type and density within each land cover category. Lidar data is ideally suited to the measurement of vegetation because it captures the shape of target objects, is spatially accurate, and can penetrate through the vegetated canopy. Airborne and ground-based Lidar were used to measure tree dimensions in a study area containing forests and woodlands representative of East Texas. The airborne lidar for

this part of the project was collected in 2004 from a 5x5 mile area near Huntsville, TX that contains a diverse assemblage of vegetation representative of East Texas, including pine and deciduous stands of various ages. This area shall hereafter be referred to as the Huntsville study area. Additional information about the Huntsville study area was collected through field work, including tree diameter at breast height (dbh), and tree species.

Land Use/Land Cover Classification of the 12km CAMx Domain

This study was created to apply existing land cover classification and supplementary data to create a LULC dataset classified to the Texas Land Classification System (TXLCS)(Texas Geographic Information Council, 1999) . The TXLCS is a vegetation-centric Anderson classification scheme that separates land use/land cover into 26 classes, including multiple classes in the Developed, Forest, Woodland, Wetland and Shrub categories (Table 2).

The grant activities description for the land use/land cover classification of the 12km CAMx domain intended to crosswalk the 2001 National Land Cover Dataset (NLCD) to the TXLCS, but after trying to crosswalk the 2001 NLCD to the TXLCS, it was determined to be insufficient (Task 3 – Identification of LULC classes that have analogues with the NLCD vs. classes without direct correlation to NLCD:). The main shortfall of crosswalking the 2001 NLCD to the TXLCS was the inadequate (or absence of) foliage typing in the 2001 NLCD for forest and shrub classes. Cold-Deciduous Forest is easily crosswalked from the 2001 NLCD to the TXLCS, but the 2001 NLCD only has one Evergreen class where the TXLCS differentiates between needle-leafed and broad-leafed evergreen. Incorporating the 2001 NLCD Canopy Cover dataset allowed for the separation of woodlands from the Cold-Deciduous Forest class, but only general Evergreen Woodland class was extracted because of the inadequate foliage typology of the 2001 NLCD needed to crosswalk to the TXLCS. With the differentiation of evergreen classes in the TXLCS, the 2001 NLCD's Mixed Forest class cannot be directly translated to the TXLCS. The Mixed Forest class in the TXLCS can be a mixture of needle-leafed evergreens and broad-leafed evergreens, but the 2001 NLCD only recognizes a mix of cold-deciduous and evergreen. Also, the 2001 NLCD only has a single class for shrubs where the TXLCS divides shrubs into five classes (Cold-Deciduous, Broad-leafed Evergreen, Needle-leafed Evergreen, Mixed, and Desert). Similar to the inadequacy for shrubs, the 2001 NLCD only has one class for woody wetlands, where the TXLCS calls for three (Riparian, Swamp Forested, and Shrub). All of the shortfalls of the 2001 NLCD for a direct crosswalk to the TXLCS are outlined in Table 3.

During his November 17th presentation at the American Society for Photogrammetry and Remote Sensing (ASPRS) and Management Association for Private Photogrammetric Surveyors (MAPPS) 2009 Specialty Conference in San Antonio, TX, Andrew Grogan used the LANDFIRE dataset in his spatial analysis model for generating composite cost surfaces to depict cross country mobility in natural terrain. The LANDFIRE dataset he described sounded like a promising alternative to the 2001 NLCD and required further investigation. After the conference, the LANDFIRE dataset was investigated. The LANDFIRE dataset was much more detailed than the TXLCS required (classified vegetation by species rather than foliage type for each vegetation type), which made it a prime candidate for crosswalking. The Existing Vegetation Type dataset within the LANDFIRE dataset was of primary importance. The Existing Vegetation Type dataset is a raster dataset with an attribute table. The attribute table has eight different classification schemes for the vegetation, consisting of Existing Vegetation Type (EVT_Name), Existing Vegetation Group (EVG_Name), Society of American Foresters and Society for Range Management (SAF_SRM), National Vegetation Classification Standard Order (NVCSORDER), National Vegetation Classification Standard Class (NVCSCLASS), National Vegetation Classification Standard Subclass (NVCSSUBCLA), System Group 1 (SYSTEMGR_1) and System Group 2 (SYSTEMGR_2). The EVT_Name classification classified the vegetation into general vegetation types by ecological zones. The EVG_Name classification added a better description of species within the EVT_Name classification in some cases. The SAF_SRM classification classified each vegetation class by species. The set of National Vegetation Classification Standard classifications (NVCSORDER, NVCSCLASS, and NVCSSUBCLA) classified by dominant vegetation type (i.e. tree-dominant, shrub-dominant), described the canopy structure (i.e., open-canopy, closed-canopy), and combined the NVCSORDER and NVCSCLASS classifications and added foliage typology (i.e. mixed

evergreen-deciduous open tree canopy), respectively. SYSTEMGR_1 indicated species (i.e., Bigtooth Maple Woodland) and SYSTEMGR_2 described the vegetation type (i.e., hardwood, conifer, etc.) The primary benefit of the LANDFIRE dataset is the need to aggregate classes that are more detailed than needed rather than split classes that are not detailed enough, as in the 2001 NLCD. After discussing the benefits of the LANDFIRE dataset over the 2001 NLCD with the TCEQ Project Manager , it was determined that the LANDFIRE dataset should be the foundation of the crosswalk rather than the 2001 NLCD. The LANDFIRE website (http://www.landfire.gov/background.php) describes the LANDFIRE project as follows:

> "LANDFIRE was also known as the Landscape Fire and Resource Management Planning Tools Project from 2004 to 2009. As a multi-partner project, it produced consistent maps and data describing vegetation, wildland fuel, and fire regimes across the United States being a shared project between the wildland fire management programs of the U.S. Department of Agriculture Forest Service and U.S. Department of the Interior."

Class Number	Class	Abbreviated Class
1	Open Water	OW
2	Developed Open Space	DOS
3	Developed Low Intensity	DL
4	Developed Medium Intensity	DM
5	Developed High Intensity	DH
6	Barren Land (Rock/Sand/Clay/Unconsolidated Shore)	BL
7	Herbaceous Natural	HN
8	Herbaceous Cultivated	HC
9	Riparian Forested Wetland	RFW
10	Swamp Forested Wetland	SFW
11	Shrub Wetland	SW
12	Herbaceous Emergent Wetland	HEW
13	Cold-Deciduous Forest	CDF
14	Broad-leafed Evergreen Forest	BEF
15	Needle-leafed Evergreen Forest	NEF
16	Mixed Forest	MF
17	Cultivated Woody Vegetation	CWV
18	Cold-Deciduous Woodland	CDW
19	Broad-leafed Evergreen Woodland	BEW
20	Needle-leafed Evergreen Woodland	NEW
21	Mixed Woodland	MW
22	Cold-Deciduous Shrub	CDS
23	Broad-leafed Evergreen Shrub	BES
24	Needle-leafed Evergreen Shrub	NES
25	Mixed Shrub	MS
26	Desert Scrub	DS

Table 2. Texas Land Classification System classes.

**********Class analogies based on definitions of NLCD 2001 and TXLCS 1999**********			
NLCD 2001 Class	**NLCD 2001 Canopy Cover**	**TXLCS Class (Phase I and II Class Numbers)**	**Comments**
11 - Open water	N/A	1 - Open water	
21 - Developed, Open Space	N/A	2 - Developed, Open Space	
22 - Developed, Low Intensity	N/A	3 - Developed, Low Intensity	
23 - Developed, Medium Intensity	N/A	4 - Developed, Medium Intensity	
24 - Developed, High Intensity	N/A	5 - Developed, High Intensity	
31 - Barren Land	N/A	6 - Barren Land (Rock/Sand/Clay/Unconsolidated Shore)	
41 - Deciduous Forest	> 50%	13 - Cold Deciduous Forest	
	< 50%	18 - Cold Deciduous Woodland	
42 - Evergreen Forest	> 50%	14 - Broad-leafed Evergreen Forest	No way to differentiate broad-leafed and needle-leafed through NLCD alone
	< 50%	19 - Broad-leafed Evergreen Woodland	
	> 50%	15 - Needle-leafed Evergreen Forest	
	< 50%	20 - Needle-leafed Evergreen Woodland	
43 - Mixed Forest	> 50%	16 - Mixed Forest	TXLCS Class could include mix of broad-leafed and needle-leafed evergreens while NLCD is only mix of deciduous and evergreen
	< 50%	21 - Mixed Woodland	
52 - Scrub/Shrub	N/A	22 - Cold-Deciduous Shrub	No way to differentiate shrub type through NLCD alone
	N/A	23 - Broad-leafed Evergreen Shrub	
	N/A	24 - Needle-leafed Evergreen Shrub	
	N/A	25 - Mixed Shrub	
	N/A	26 - Desert Scrub	
71 - Grassland/Herbaceous	N/A	7 - Herbaceous Natural	
81 - Pasture/Hay	N/A	8 - Herbaceous Cultivated	
82 - Cultivated Crops	N/A		
	N/A	17 - Cultivated Woody Vegetation	No way to separate from other woody vegetation through NLCD
90 - Woody Wetlands	N/A	9 - Riparian Forested Wetland	
	N/A	10 - Swamp Forested Wetland	
	N/A	11 - Shrub Wetland	No way to separate Woody Wetland types
95 - Emergent Herbaceous Wetland	N/A	12 - Herbaceous Emergent Wetland	

Table 3. Defines the NLCD classes analogous to the TXLCS.

The LANDFIRE dataset covers the entire United States and is the foundation for the TCEQ CAMx 12km land cover dataset (TCEQLC_2010). The LANDFIRE dataset differentiates between individual species and species mixes, which is more detailed than the TXLCS requires, so the LANDFIRE classes were aggregated to the TXLCS classes. To aggregate the LANDFIRE classes to the TXLCS class level, the LANDFIRE data were analyzed by SYSTEMGR_1, SYSTEMGR_2 and NVCSSUBCLA (Task 3.4 – Conversion Routines for converting NLCD to TX LULC). SYSTEMGR_1 was used to directly crosswalk the non-vegetated classes to the TXLCS. SYSTEMGR_2 and NVCSSUBCLA were used to translate the vegetated LANDFIRE classes to the TXLCS by using the SYSTEMGR_2 classification to classify to the general type (Hardwood went to Cold-Deciduous, Conifer to Needle-leafed Evergreen, and Shrubland went to Shrub). The SYSTEMGR_2 classified Hardwoods were then investigated according to their NVCSSUBCLA classification. When the SYSTEMGR_2 classification matched the NVCSSUBCLA (Hardwood and Deciduous), the TXLCS class stayed Cold-Deciduous; however, when the two did not match (Hardwood and Evergreen), the TXLCS class changed to Broad-leafed Evergreen. The classes earmarked as shrublands were then classified according to their species found in SYSTEMGR_1. This process crosswalked all the LANDFIRE classes to the TXLCS level except separating Forest from Woodland. While NVCSCLASS distinction between open- and closed-tree canopy could have been used to separate Forest and Woodland classes, the terms open and closed tree canopy were too ambiguous to use as a rule. To overcome this shortfall the LANDFIRE data was combined with the NLCD Canopy Cover to differentiate the Forest and Woodland classes (using 50% canopy cover as the dividing line), completing the crosswalk foundation.

The final classification is an amalgamation of 5 GIS layers (4 used directly and 1 indirectly) and 1 source of remote sensing data (Task 3.1 – Identification and integration of ancillary GIS data:). The list of GIS data includes:

LANDFIRE
NLCD Canopy Cover
TPWD
INEGI's Land Use and Vegetation and
INEGI's Hydrographic Features

The Texas Parks and Wildlife Department (TPWD) Texas Ecological Systems Classification Project (http://www.tpwd.state.tx.us/landwater/land/maps/gis/tescp/index.phtml) was used to enhance the TCEQLC_2010. It provided land cover resolved at 10m and finer vegetation classes for East and Central Texas, making it easier to separate mixed classes into their individual TXLCS class. The TPWD data was overlaid onto the TCEQLC_2010 foundation after aggregating classes to the TXLCS level and separating Forest and Woodland classes with the NLCD Canopy Cover dataset.

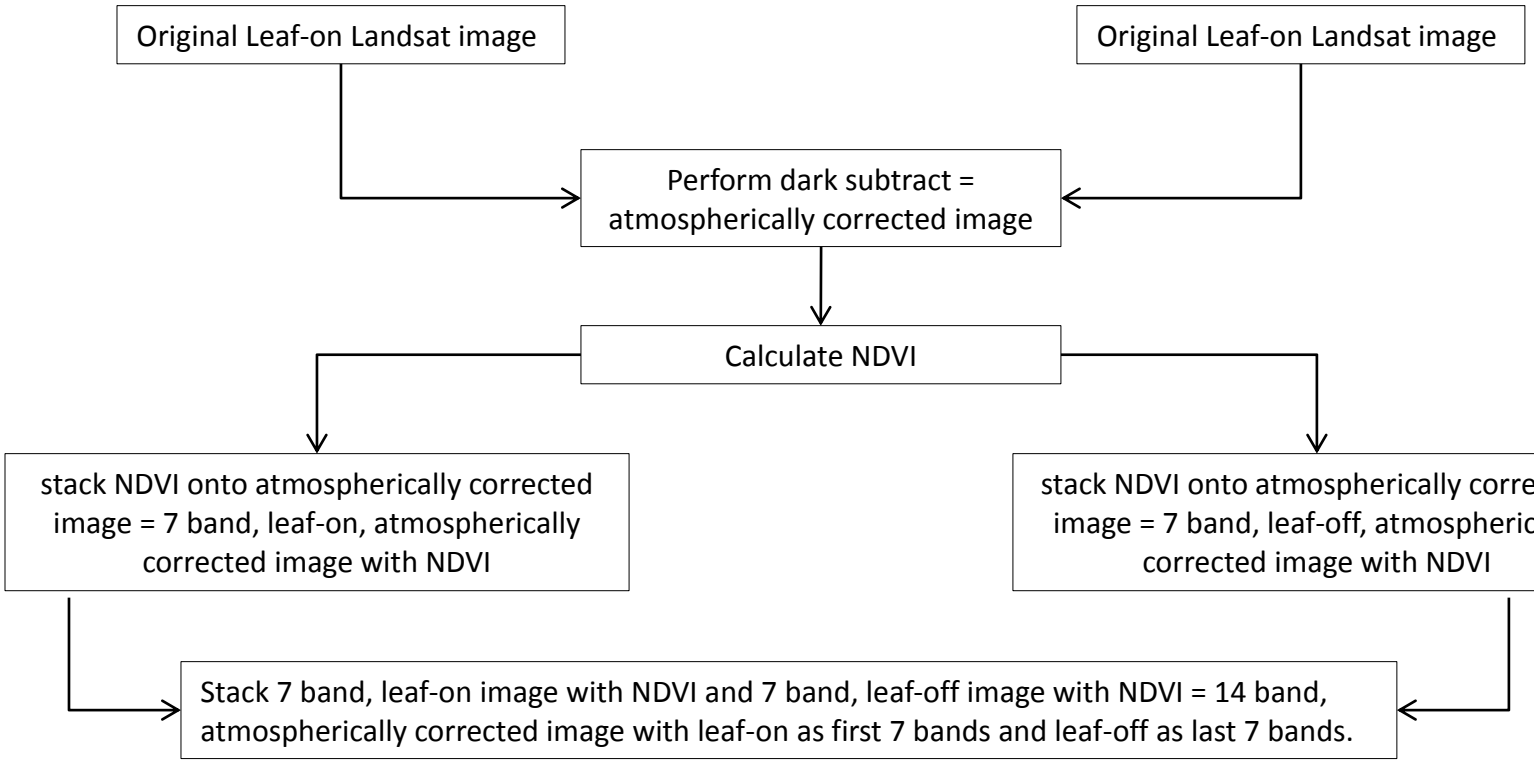
While the LANDFIRE dataset covers the entire United States, the 12km CAMx grid extends into northern Mexico and required classification of LandSat imagery. Two datasets created by the Mexican Instituto Nacional de Estadistica y Geografia, National Institute of Statistics and Geography, (INEGI) were used to supplement the LandSat image classification. The 250 meter resolution Land Use and Vegetation (Uso de Suelo y Vegetacion) was used to guide the classification as to what land cover was in the area because site visits were not possible. Hydrographic Features (rasgos hidrograficos) data for water bodies were used to identify the water class for the portion of the study area in Mexico.

To classify the entire 12km CAMx extent, LandSat imagery had to be classified (Task 3.3 – Supplemental Imagery Classification:). LandSat imagery was chosen because it has global coverage, is free, and matches the resolution of the LANDFIRE data. At least two images were obtained for each LandSat scene, one leaf-on and one leaf-off. The leaf-on imagery indicates where all vegetation is, while the leaf-off only shows the evergreen and winter vegetation. This is a simple way to differentiate deciduous from evergreen vegetation. Table 4 lists the 82 LandSat images that were classified to create the TCEQLC_2010 dataset. Dark subtract was used for atmospheric correction by subtracting the lowest pixel value making up 1% of the non-zero pixel values for each band. After performing the dark subtract, a normalized difference vegetation index (NDVI) layer was calculated for each image using Equation 1, below. NDVI is used to quickly identify areas of green vegetation. The NDVI layer was then stacked onto the original image creating a 7

band image. Same scene images were then stacked to create a 14 band image with the leaf-on image comprising the first 7 bands and leaf-off being the last 7 bands, illustrated in following flow chart.

$$NDVI = \frac{NIR - R}{NIR + R}$$

Equation 1. Equation for calculating NDVI where NIR is the near infrared band and R is the red band.



For each LandSat path/row:

Path	Row	Year	Day of Year
26	42	2003	129
26	42	2003	353
26	42	2005	102
26	43	2001	203
26	43	2002	14
26	43	2005	102
27	42	2000	96

27	42	2001	290
27	42	2002	5
27	42	2004	27
27	43	2000	96
27	43	2001	290
27	43	2001	50
27	43	2002	53
28	40	2001	297
28	40	2003	47
28	41	2001	297
28	41	2002	268
28	41	2003	351
28	42	2001	329
28	42	2001	361
28	42	2002	268
28	43	2001	329
28	43	2003	335
29	40	2001	288
29	40	2002	51
29	41	2001	288
29	41	2003	54
Path	Row	Year	Day of Year
29	42	2002	291
29	42	2003	54
29	43	2002	291
29	43	2003	54
30	39	2001	295
30	39	2002	346
30	40	2001	295

30	40	2002	362
30	41	2001	295
30	41	2003	29
30	42	2001	7
30	42	2001	295
30	42	2003	333
30	43	2001	7
30	43	2001	279
31	39	2001	14
31	39	2001	270
31	40	2001	270
31	40	2002	17
31	41	2002	17
31	41	2002	241
31	42	2002	17
31	42	2002	241
31	42	2002	273
31	43	2002	1
31	43	2002	273
32	39	2001	53
32	39	2002	8
Path	Row	Year	Day of Year
32	39	2003	139
32	40	2002	8
32	40	2004	254
32	41	2002	8
32	41	2004	254
32	42	2001	293
32	42	2003	27

32	43	2001	21
32	43	2004	286
33	38	2003	194
33	38	2003	34
33	39	2003	194
33	39	2003	18
33	40	2001	156
33	40	2001	252
33	40	2003	18
33	41	2001	156
33	41	2003	130
33	41	2005	7
33	42	2003	2
33	42	2004	165
33	42	2004	309

Table 4. List of LandSat scenes used for supplemental image classification.

Each of the 14 band images were imported into the Definiens object-based classification software package for classification. Each image was segmented to group similar contiguous pixels and segments were then classified. The INEGI Land Use and Vegetation data, along with Google Earth, were used to guide the selection of training areas. Training areas were selected from the areas of overlap between images as much as possible to minimize incongruence between neighboring images which can cause seams in the final dataset. All classes were considered in the classification except the four Urban classes and the Water class. The Urban classes were not considered because many of the urban areas were too spectrally similar to the Shrub classes, specifically Desert Scrub, and the INEGI Land Use and Vegetation dataset contained an Urban class that could be overlaid onto the classification. The Water class was not used because shadows in mountainous regions would be classified as Water, causing a gross overestimation, and the INEGI Hydrographic Features dataset contained everything needed.

The final classification was a combination of the classification created from the LANDFIRE, TPWD and NLCD canopy cover datasets, and the classified LandSat images and INEGI data (Task 3.5 – Final Classification and Shapefiles:). First, the data covering the portion of the study area in Mexico was assembled. Then, the classifications of the LandSat imagery were mosaicked to create a base layer. The Urban areas from the INEGI Land Use and Vegetation dataset were to be overlaid onto the base layer next. The Urban class was split into three classes based on the size: Low, Medium, and High Intensity Developed. Intensity classes were determined by which quartile the Urban class fell into. Urban areas whose area was within the first two quartiles were designated as Low Intensity, with Medium and High Intensity classes designated to Urban areas with sizes in the third and fourth quartiles, respectively. With the Urban areas in place, the

Water class was overlaid. Next the U.S. and Mexico classifications were mosaicked, with the U.S. classification taking precedence in areas of overlap, creating the TCEQLC_2010 dataset (Figure 2).

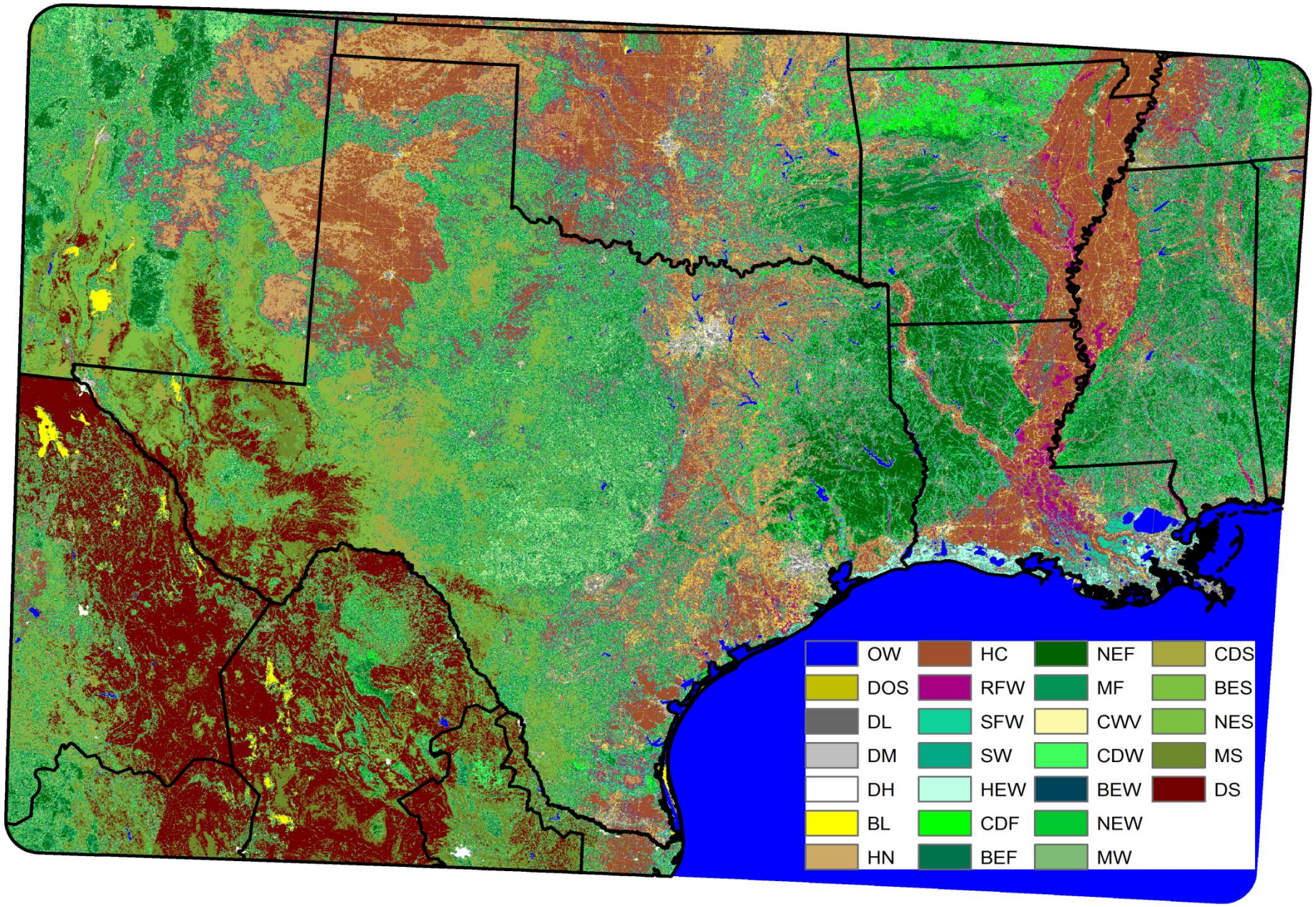


Figure 2. Final classification of the 12km CAMx domain, TCEQLC_2010.

The classification was then subset by county/parish for areas within the U.S. and by state for areas within Mexico and converted to a shapefile. To automatically use the TCEQLC_2010 dataset and a shapefile consisting of all the counties/parishes and Mexican states, a python script was written. A Python script was written to automatically subset the TCEQLC_2010 raster dataset by U.S. county/parish and Mexican state and converts the raster subsets to shapefiles. The code creates a folder for Mexico with a shapefile for each Mexican state and a folder for USA with folders for each state containing shapefiles for each county/parish. The code can be found in APPENDIX B.

An accuracy assessment was not conducted on the finished product because accuracy could not be reliably determined for Mexico due to a lack of ancillary data with high enough resolution and detailed classes or ~~in situ~~in-situ data. However, accuracy assessments were conducted for super zones in the LandFire data. The lowest accuracy reported for a super zone within the study area was 44%. Although this is low, it was calculated at a more detailed level than needed for the TXLCS. When aggregated to the TXLCS level, the accuracy should be much higher. The results of each accuracy assessment within the study area are summarized in Table 5 and can be found, in full, at:

http://www.landfire.gov/dp_quality_assessment.php[1]

Super Zone	Overall Agreement
Southwest	44.0%
South_Central_West	59.1%
Southern Appalachians	66.2%
Northern Plains	68.3%
South_Central_East	84.0%

Table 5. Results of regional accuracy assessments of the LANDFIRE dataset.

This crosswalk of LANDFIRE data to TXLCS and LandSat image classification provides a LULC classification that has not been available before for air quality modeling. This classification should provide a better base for air quality modeling because it has a detailed classification of the portion of the 12km CAMx domain in Mexico. Previous models have used more general classes than the TXLCS. The TCEQLC_2010 dataset is intended to be updated when new LULC datasets and updates to the LANDFIRE dataset become available.

[1] Follow the link and scroll down to the map. To download the accuracy tables, click the super zone of interest in the map.

Deriving Individual Tree Characteristics from Airborne Lidar Data

While the TCEQLC_2010 allows for air quality modeling on a large scale based on emission variables estimated based on the land cover, lidar can be used to better estimate the emissions at the local level. Lidar is an active remote sensing technology that can measure the distance to an object by illuminating an object with light in the form of a laser pulse and timing the duration it takes the pulse to return to the sensor (Popescu et al., 2004). In forested areas, lidar data can be used to derive tree height, diameter at breast height (dbh), crown base height (CBH), crown diameter, and biomass. Lidar is known to not intercept the exact treetop and rather the shoulder of the tree leading to a slight underestimation of tree height by around 1-2 meters. Knowing all of these variables can help to calibrate the air quality models for the area.

The study site was a 47km² area located just east of Huntsville, TX, in the Sam Houston National Forest (Figure 3). The area is dominated by loblolly pines, made up of both plantations and natural stands, with some hardwoods along creek beds. The area also contains some urban interface and open fields. The lidar data was acquired in March of 2004, leaf-off season. Sixty-two plots were surveyed and measurements were taken for total tree height, crown diameter, height to crown base, and dbh for each tree as well as tree location (Task 3.2 – Field Work:).

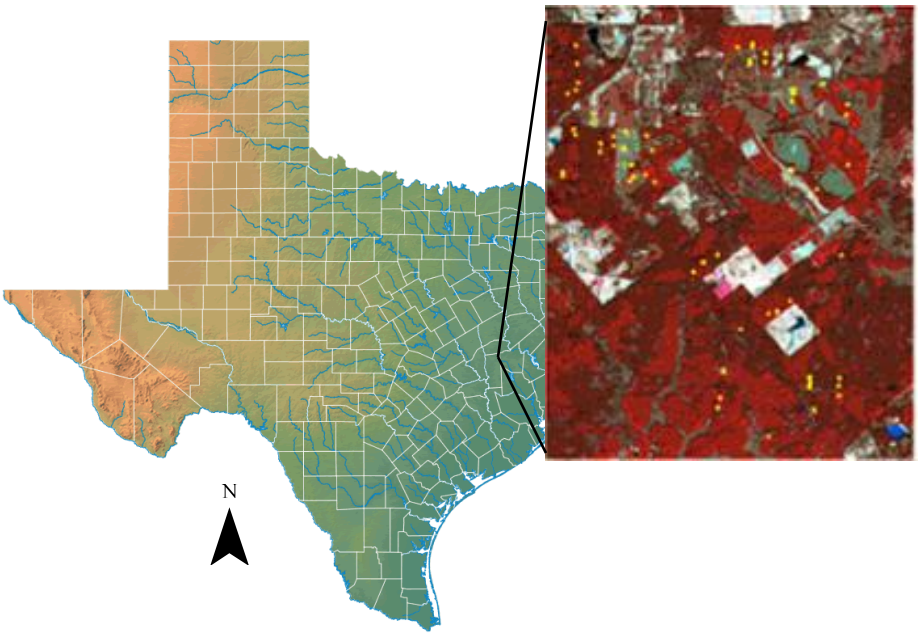


Figure 3. Study area with plot locations indicated by yellow dots.

The first step taken to derive the variables listed above (Task 4.4 – Maps of individual tree crown dimensions:) with the lidar data was to create a Canopy Height Model (CHM). A CHM (Task 4.0 – Canopy Height Model) is a gridded surface that depicts the height above ground of the highest lidar return within each grid cell. A bare earth model and digital surface model (DSM) were created to determine ground level and uppermost return, respectively. The bare earth model was subtracted from the DSM to create a relative height to ground dataset. The CHM was created at 0.5m resolution. At 0.5m, there were holes in the CHM due to no lidar hits within the grid cell. A 3x3 moving window

maximum filter was applied to remove the holes by replacing the center cell with the maximum value within the 9 cells. Figure 4 depicts the height above ground at 0.5m resolution.

Figure 4. 0.5 meter resolution canopy height model of the Huntsville, TX study area.

Once the CHM was created, TreeVaW (Popescu and Wynne, 2004) was used to identify each tree and output XY location, tree height, and crown diameter. TreeVaw uses a local maxima filter with a variable, circular window. The size of the window is determined by the height of the tree, with the assumption that taller trees have wider crowns. Through regression of the field-measured trees, dbh was estimated with an RMSE of 4.5cm using the relationship of tree height and crown diameter to dbh (Popescu, 2007).

Zhao et al., (2009) estimated individual tree CBH from lidar height bin products and used it to estimate individual tree biomass. Height bins were created by "slicing" the point cloud horizontally at a vertical interval. The height bins created for estimating CBH used voxel dimensions of 0.5 x 0.5 x 1.0 m (x, y, z), with up to 31 height bins starting at 0.0 m. For each TreeVaw identified tree, the height bins were cut vertically, centered at the tree location with a diameter equal to the TreeVaw identified crown width. The cylinder of height bins was used to create a vertical frequency profile, with CBH being identified as the height at which an abrupt drop in the vertical profile occurs. Popescu and Zhao (2008) used this approach and achieved an RMSE less than 2.0m.

The dbh is the most common variable used to estimate biomass (Task 4.1 – Above-ground Biomass Map:) through the use of allometric equations; however, lidar data cannot be used to directly measure dbh. Two regression analyses (one for pines and one for deciduous) were used to estimate dbh from the lidar-derived height, crown width and CBH by relating the field-measured dbh to the lidar-derived tree dimension variables. The estimated dbh was then plugged into general allometric equations to calculate total above ground biomass for each identified tree. The above ground biomass estimate of each tree was then apportioned into stem and foliage and branch biomass according to ratio equations. The stem biomass was assigned to the pixel at the tree location and the foliage and branch biomass was distributed with a 2-D Gaussian distribution over the pixels covered by the crown. Below is the resulting biomass map.

Figure 5. Image of the distribution of the biomass map.

Surface Roughness

Surface roughness (Task 4.2 – Surface Roughness:) is a variable used to model turbulence. This report investigates the effectiveness of using the Shuttle Radar Topography Mission (SRTM) data along with the National Elevation Dataset (NED) to create a CHM for use in estimating surface roughness by comparing the SRTM-derived CHM to a lidar-derived CHM. Surface roughness was calculated to coincide with the 4km grid used for modeling by calculating the variance within each 4km grid cell.

The SRTM data was collected between February 11, 2000 and February 22, 2000 and covered the land surface of earth between 60° North latitude and 56° South latitude, roughly 80% of all earth land surface. For North America, the data is available at 30m resolution. The Jet Propulsion Lab (JPL) has reported accuracies for the SRTM data as 12.6m, 9.0m, and 7.0m for absolute geolocation error, absolute height error, and relative height error, respectively (Farr et al., 2007).

The SRTM and NED data were projected from Global Coordinate System to UTM Zone 15N in the NAD83 datum. The NED was then subtracted from the SRTM data to create a CHM. The CHM was then resampled to 0.5m to match the lidar-derived CHM for the following reasons: (1) The lidar-derived CHM had to be cleaned to remove obvious errors in the data, such as points above 50m and below -5m. Even though there were less than twenty of these errors, the following procedures would have drastically increased the errors above 50m. As mentioned earlier, the lidar-derived CHM also had holes in it (because of high resolution) where there were gaps in the tree canopy. These holes would bias the surface roughness value by inflating the variance. To mimic the SRTM-derived CHM's 30m resolution and plug the holes in the CHM, a maximum filter was passed over the lidar-derived CHM. The maximum filter replaces the center pixel of the window with the highest value within the window. A window of 59 pixels x 59 pixels (29.5m x 29.5m) was used to replicate the SRTM-derived CHM resolution. Although the filtering window was very large, the pixel size remained at 0.5m. (2) Resampling the lidar-derived CHM to 30m presents a problem because it could possibly change the height values more than the maximum filter, reducing the accuracy of the CHM. There are two options for resampling in ENVI, pixel aggregate and nearest neighbor. Pixel aggregate simply averages the pixels that will go into the 30m pixel, intrinsically lowering the value of the CHM. Nearest neighbor uses the nearest pixel value as the output value, possibly selecting a lower value than representative of the full 30m pixel. By resampling the SRTM-derived CHM to 0.5m using nearest neighbor, these errors were avoided.

To compare the two CHMs, the SRTM-derived CHM was subtracted from the cleaned, max-filtered lidar-derived CHM. Also, the two CHMs were imported into ArcMap for zonal statistics to be run at the 4km grid cell size to compare variance/surface roughness.

The table (Table 6) below shows the min and max of the SRTM-derived, lidar-derived, and max filtered lidar-derived CHMs (full stats files found in statistics folder).

CHM	Min	Max
SRTM	-6.087448	26.828003
Lidar	-13.858839	135.996715
Cleaned, Max Filtered Lidar	-0.242327	49.955784

Table 6. Table of minimum and maximum heights for each dataset.

The following figures show the CHMs from the SRTM (Figure 6), original lidar (Figure 7), and cleaned, max filtered lidar (Figure 8) (tiff images in the CHM folder). The first three images are shown in ArcMap using the same value range of -14 to 136 to match the original lidar-derived CHM, and the last two figures (Figure 9 and Figure 10) use a value range of -7 to 50 to match the full range of the SRTM and cleaned, max filtered lidar-derived CHMs.

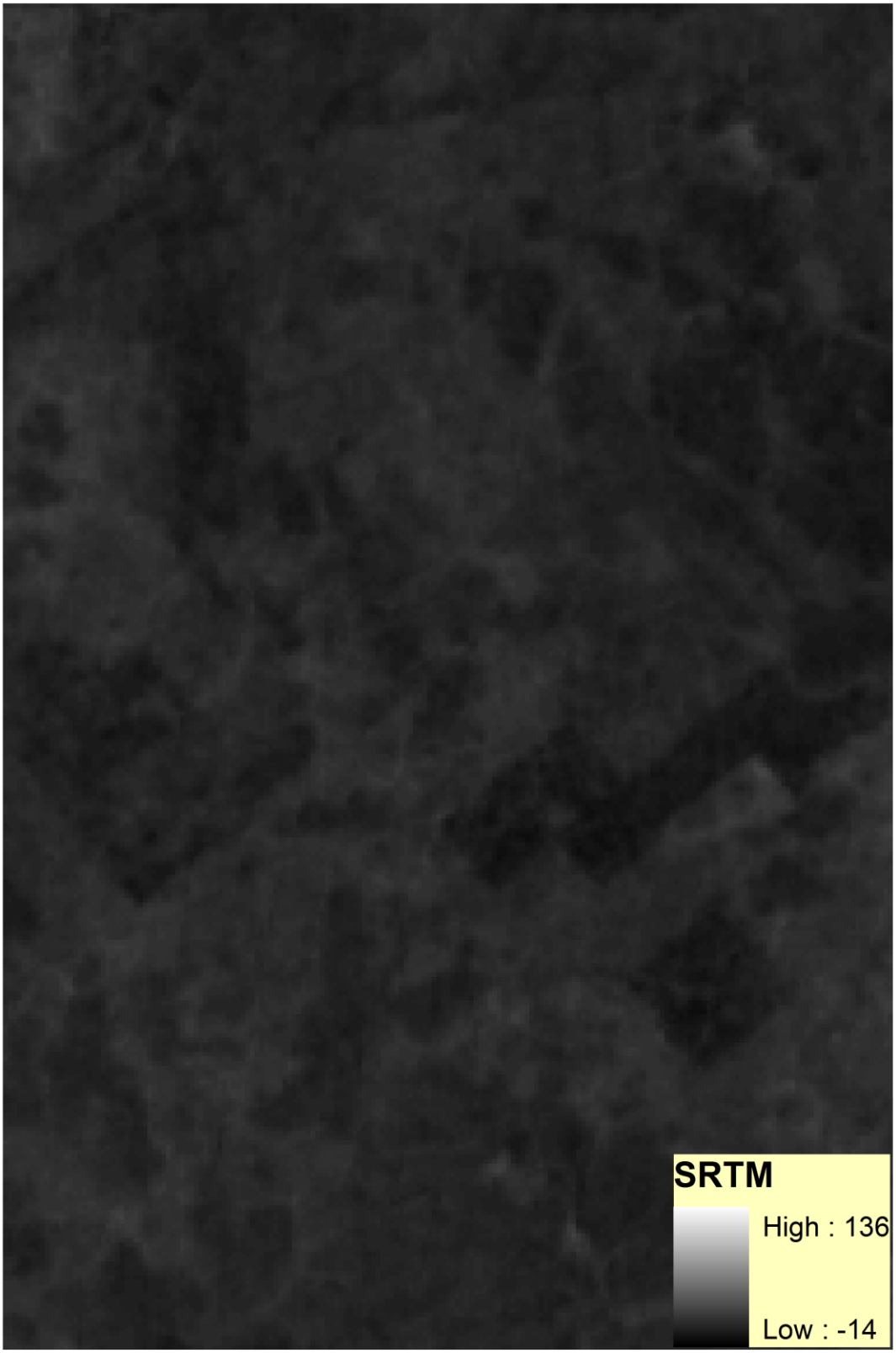


Figure 6. SRTM canopy height model using the full range of values.

Figure 7. Original Lidar canopy height model using the full range of values.

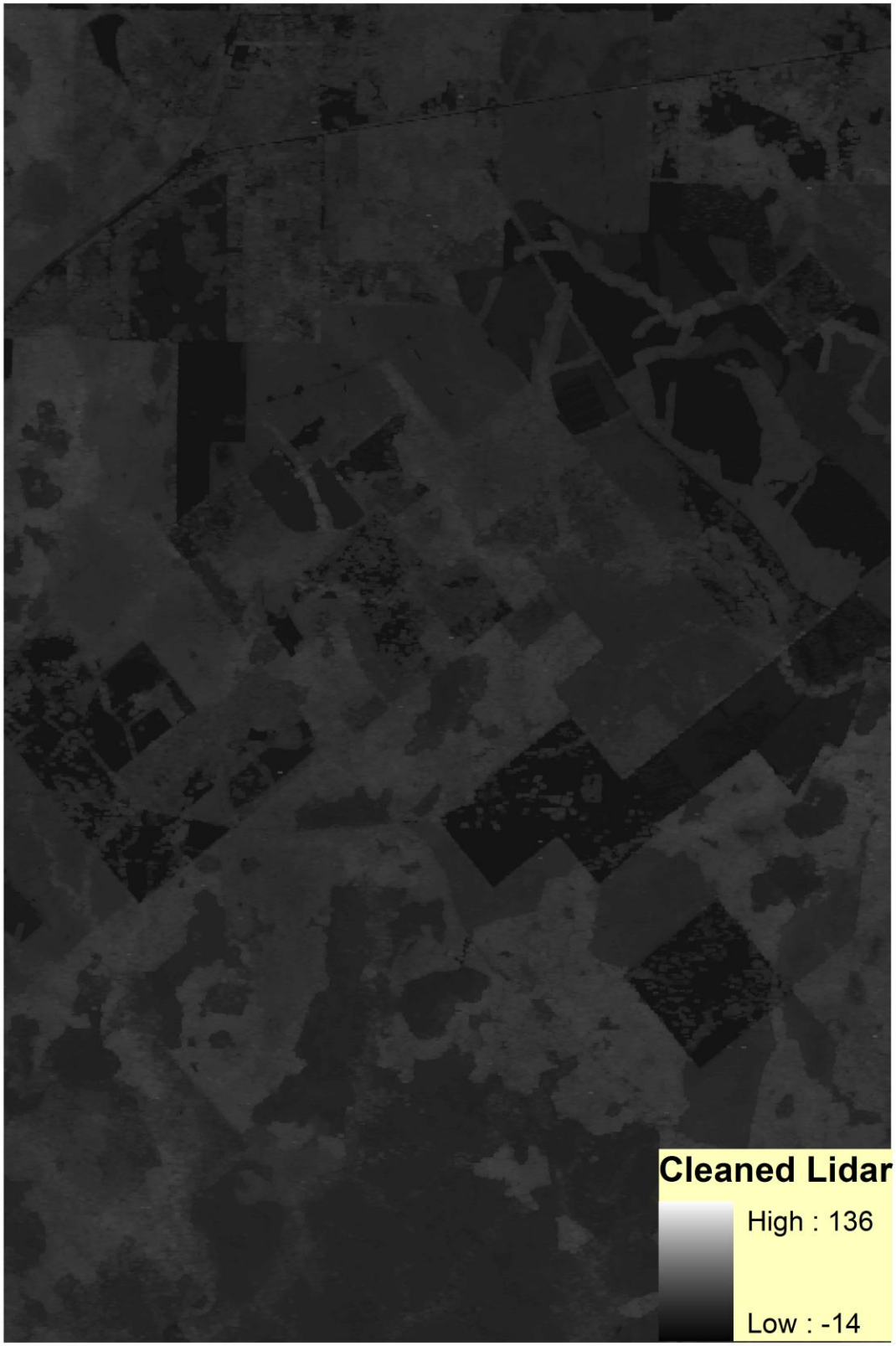


Figure 8. Cleaned Lidar canopy height model using the full range of values.

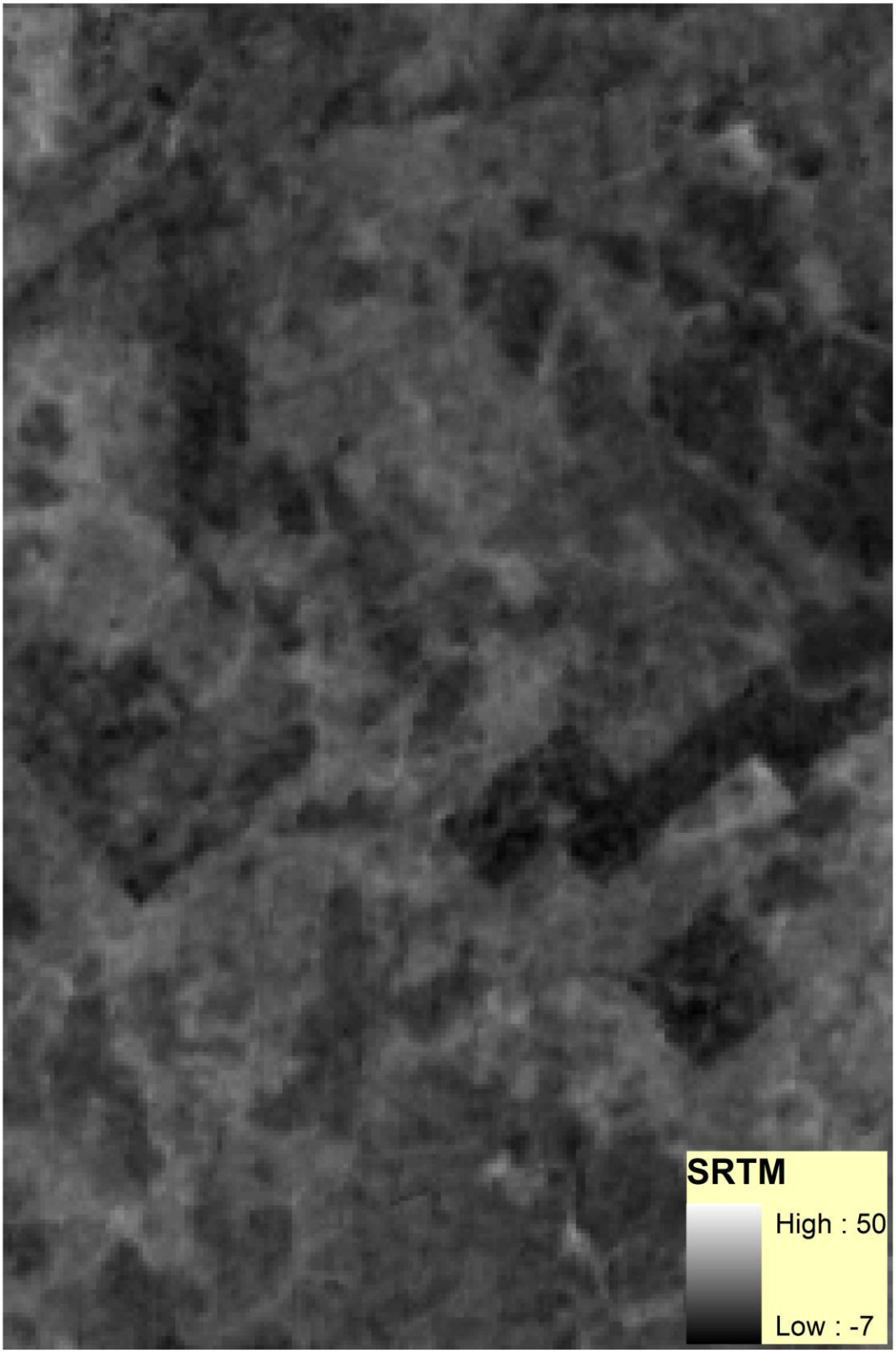


Figure 9. SRTM canopy height model using the cleaned lidar dataset's range of values.

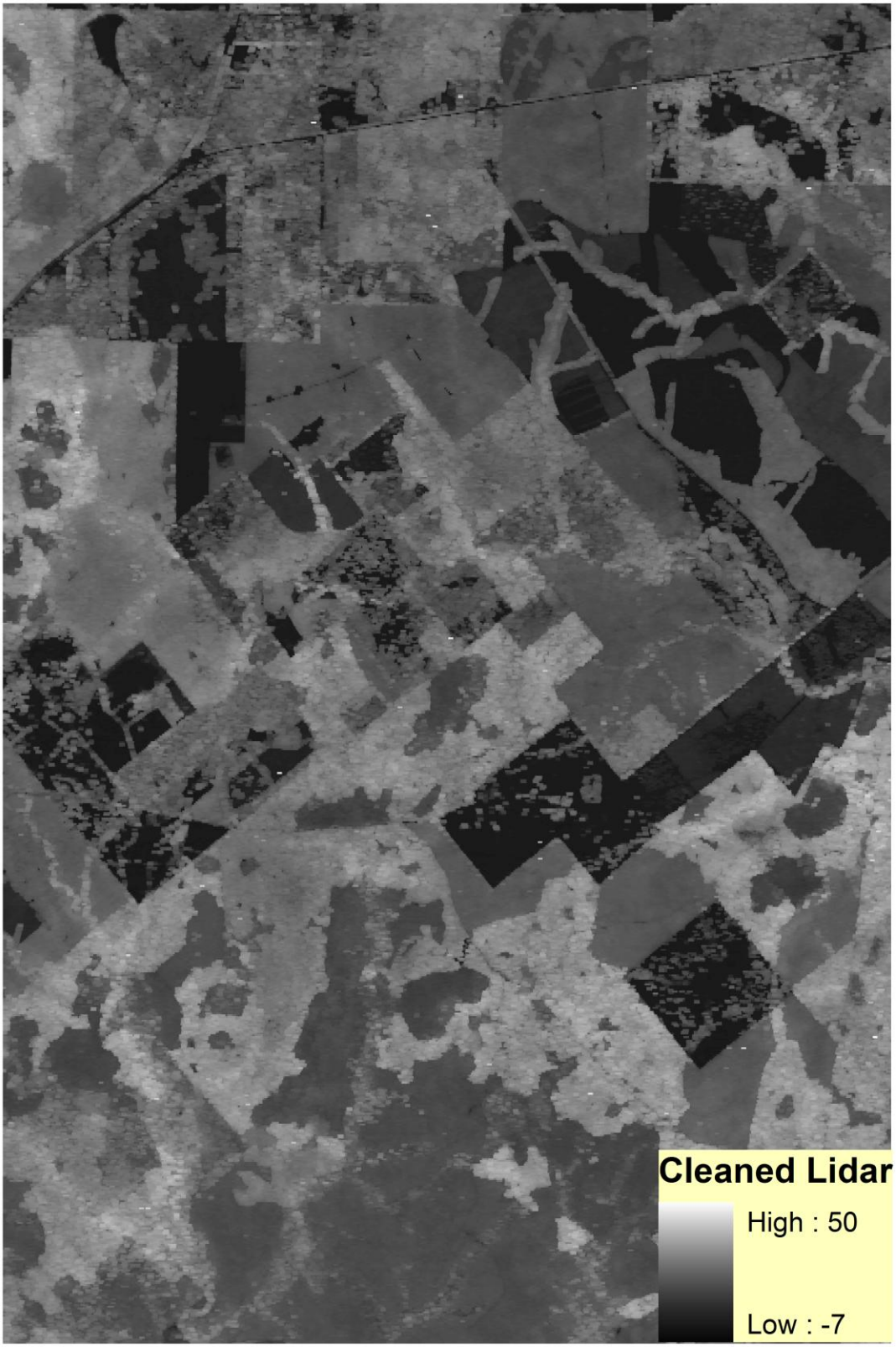


Figure 10. Cleaned Lidar canopy height model using the cleaned lidar dataset's range of values.

Figure 11 shows the result of subtracting the SRTM-derived CHM from the cleaned, max filtered lidar-derived CHM. Positive numbers show the lidar-derived CHM to be higher, while negative values show where the SRTM-derived CHM is greater. This process resulted in a minimum of -22 meters and maximum of 45 meters.

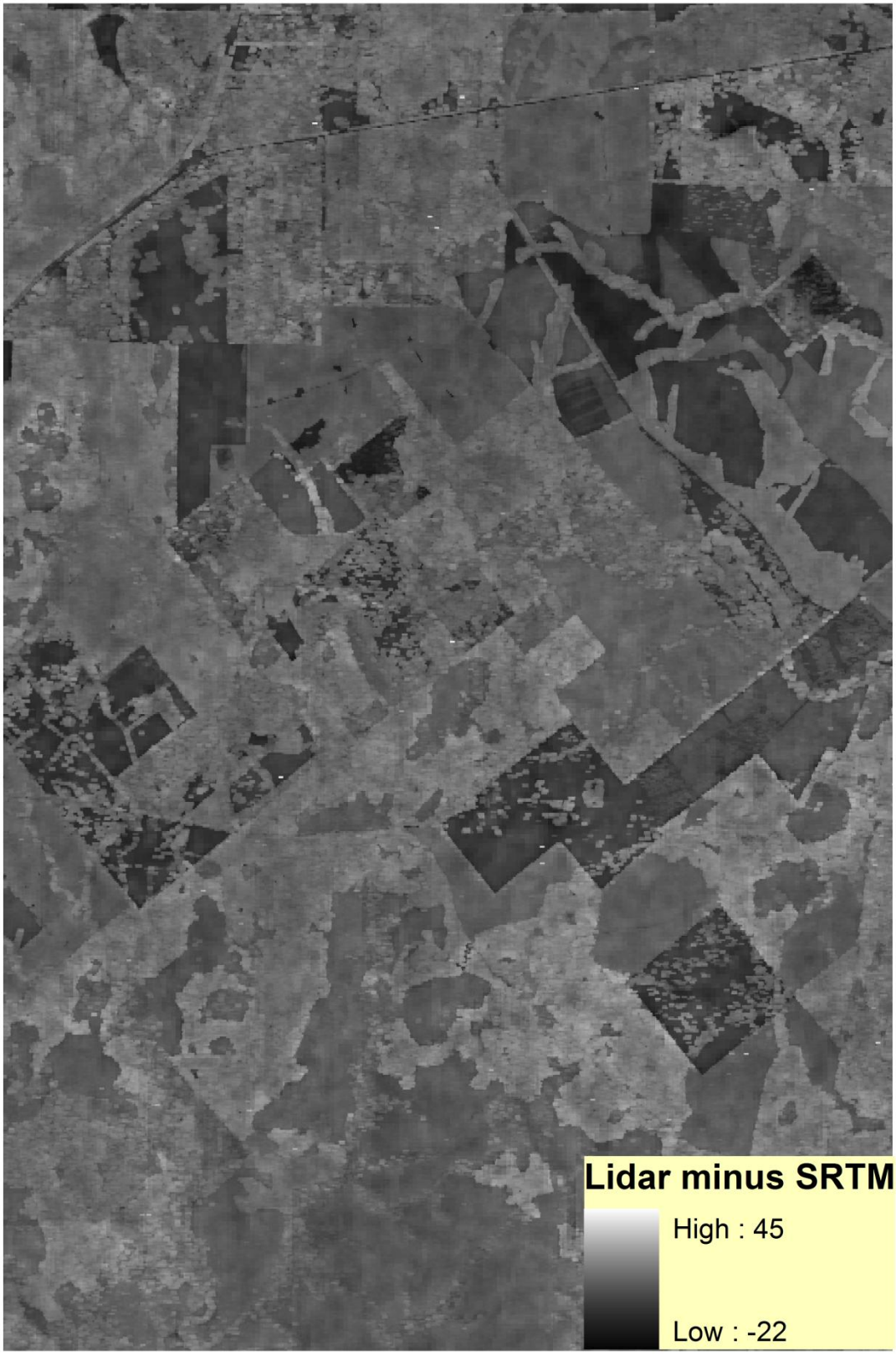


Figure 11. Lidar derived CHM minus SRTM derived CHM.

The following figures (Figure 12 - Figure 15) show zonal statistics of standard deviation and variance based on the SRTM-derived and cleaned, max filtered lidar-derived CHMs.

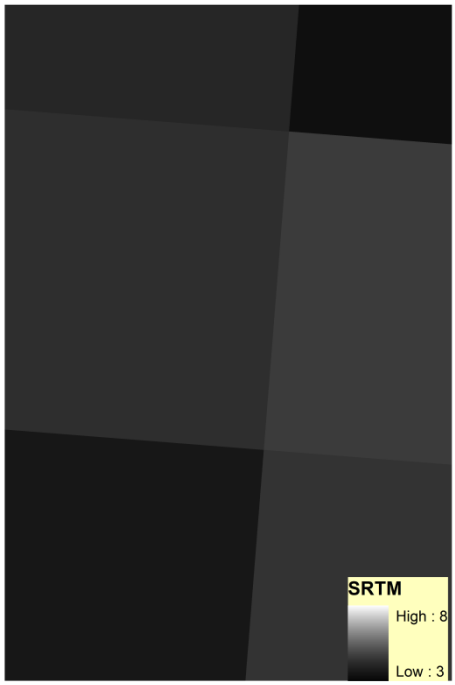


Figure 12. Standard deviation of the SRTM data at 4km resolution.

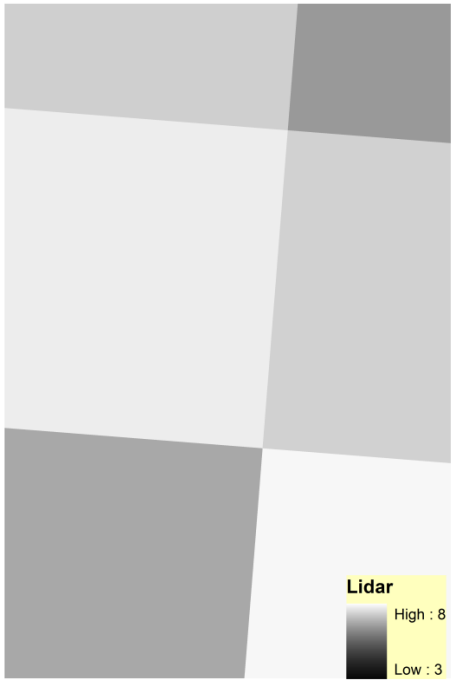


Figure 13. Standard deviation of the lidar data at 4km resolution.

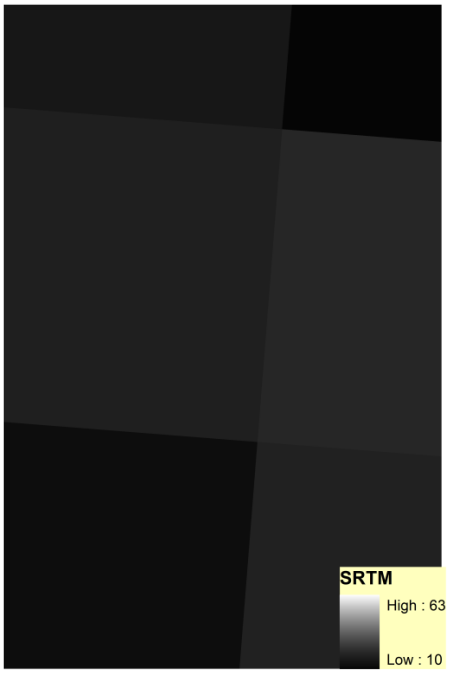


Figure 14. Variance of the SRTM data at 4km resolution.

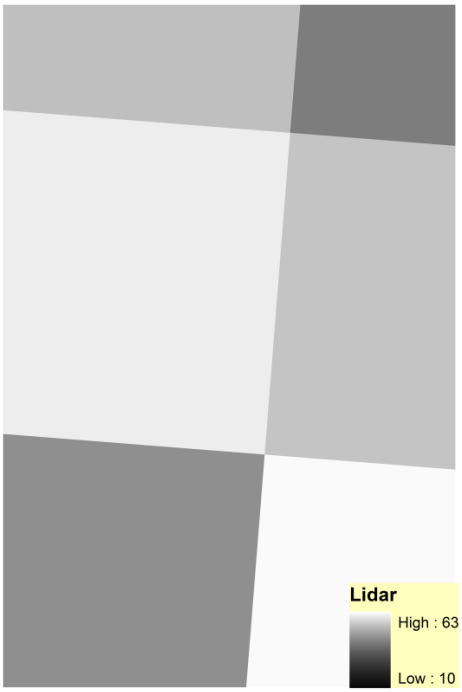


Figure 15. Variance of the SRTM data at 4km resolution.

While the SRTM and lidar-derived CHMs are different, there are a few points to keep in mind about the two data sets. First, the SRTM data was collected 4 years prior to the lidar data collection, which means the SRTM-derived CHM should be lower than the lidar-derived CHM because the trees had four more years to grow. Second, the accuracy of the two sensors also needs to be taken into account. The SRTM sensor accuracy was reported earlier being 12.6m, 9.0m, and 7.0m for absolute geolocation error, absolute height error, and relative height error, respectively, while the lidar sensor has accuracies of 40cm and 15cm for horizontal and vertical errors, respectively. Lastly, the SRTM data was collected at a 30m resolution while the lidar data had multiple returns per square meter, leading to more detail in the lidar dataset.

This is most evident in fields with few trees, where the lidar-derived CHM easily separates trees from ground while the SRTM data overestimates the area covered by trees.

The standard deviation and variance images cannot be used to directly compare the two CHMs. The values are much higher for both standard deviation and variance for the lidar-derived CHM than the SRTM-derived CHM because of the reasons mentioned earlier. They can, however, be compared for pattern. The standard deviation and variance images show a similar pattern in most of the image areas. The eastern central area is in a different order of least to greatest when comparing the two CHMs. This is most likely due to the increased detail available through lidar which clearly shows all of the fields in this area while they are not as noticeable in the SRTM data. Some of the fields could also be clear-cut areas that happened after the SRTM data acquisition.

Overall, the SRTM data can be used in conjunction with NED to create large area (regional, statewide, or national) CHMs. The SRTM is better suited for large scale modeling than for this study area and should perform well for the modeling conducted at TCEQ.

Comparison of Terrestrial Lidar to Airborne LiDAR and Ground Measurements

Terrestrial lidar is a burgeoning area for lidar because it allows for survey grade detail to be achieved. The Leica ScanStation 2 used in this study has reported accuracies of within 6mm for position and 4mm for distance compared to typical airborne lidar systems having horizontal accuracies of about 40cm and vertical of about 15cm. A comparison between the height profile of the airborne lidar and ground based lidar is shown in Figure 16 and Figure 17, below (Task 4.5 – Comparison of ground and airborne LiDAR measurements compared). While airborne lidar is known to hit the shoulder of tree crowns and not the exact tree tops, the airborne lidar still shows a taller profile. This could be due to the ground based lidar having only a single return and the beam being intercepted before reaching the tree top. The ground based lidar beam also had a high probability of not reaching the tree top because of the scan angle. If the ground based lidar had been stationed further away from the plot and had a clear view to the tree tops, it would either match the airborne lidar height profile or be slightly taller. The airborne lidar may provide a more consistent height profile in general, but the ground based lidar can more accurately measure CBH and directly measure dbh, but for a much smaller areas.

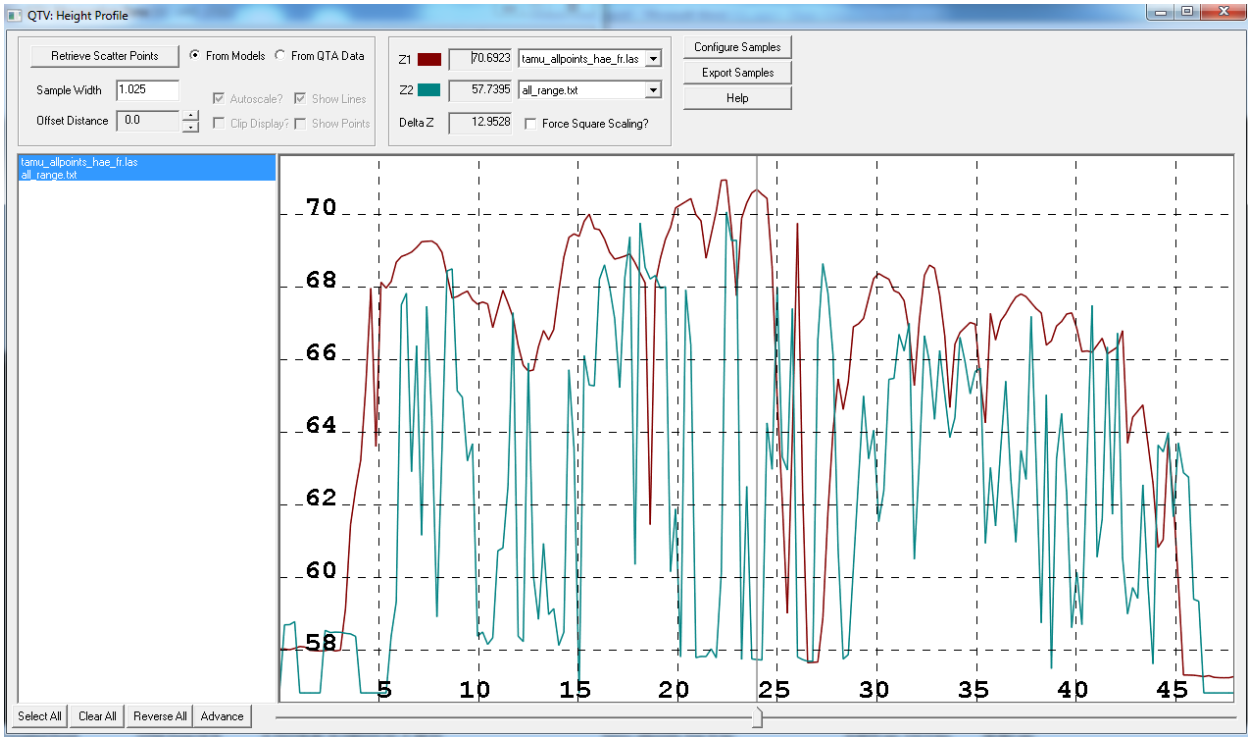


Figure 16. North-South profile comparison of plot scanned by airborne (red) and terrestrial (blue) lidar.

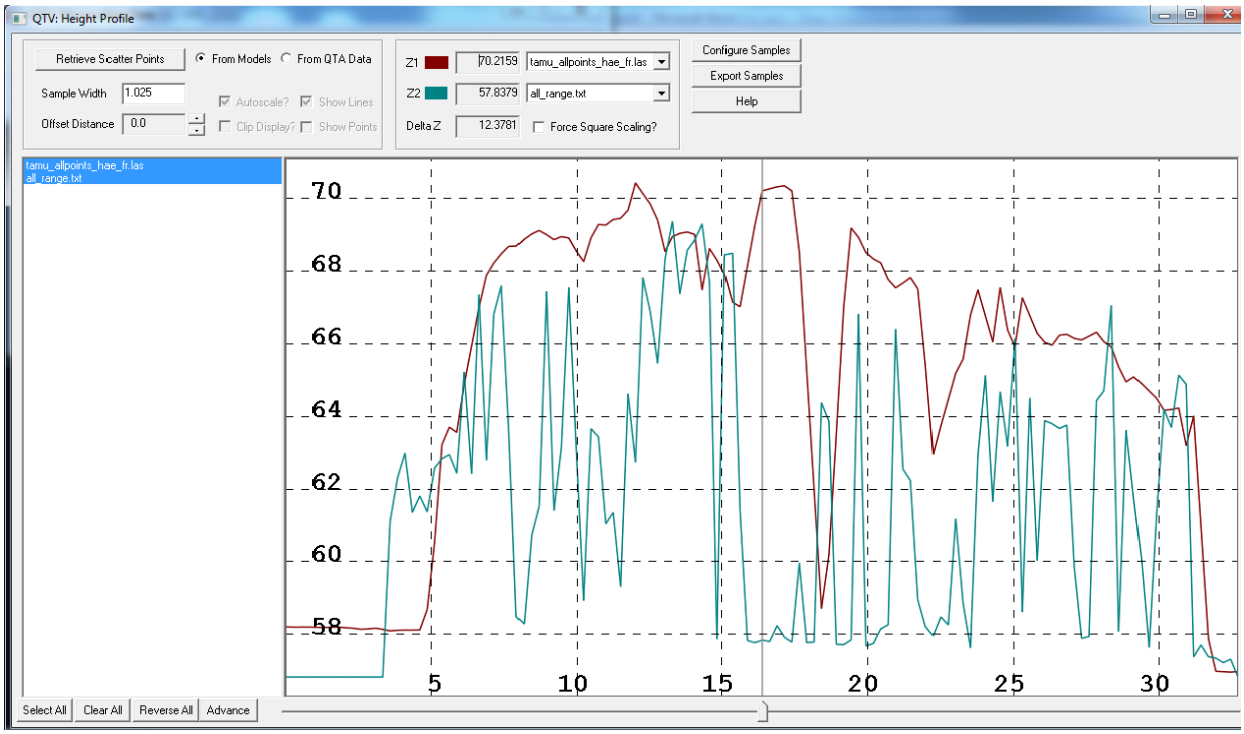


Figure 17. West-East profile comparison of plot scanned by airborne (red) and terrestrial (blue) lidar.

When compared to ground measurements (Task 4.3 – Ground-based LiDAR field measurements:), CBH estimated by using ground-based lidar data is the most reliable at this point in time. Terrestrial lidar matches the ground measurements of CBH best, because the scanner is set up to scan the underside of the canopy without having any obstructions. As mentioned earlier, the measurement of total tree height is not as accurate because of the likelihood of the beam being intercepted before reaching the tree top. Measuring dbh with terrestrial lidar is currently somewhat difficult. To measure dbh with terrestrial lidar, the user manually uses the distance tool in the software to get dbh. Automated methods for data processing to extract dbh measurements have not yet been developed by our research group. However, terrestrial lidar allows for direct measurement of dbh while airborne must rely on relationships to tree height and crown diameter. Once the measurement of dbh is automated, it will be a valuable asset to field crews because it can capture the data and eliminate the need for site revisits by other crews to check their work. Instead, they can just check the lidar data, saving time and money. Because terrestrial lidar can directly measure CBH, its biggest benefit could be in the form of a sampling tool for calibrating airborne data for CBH estimation.

Canopy Bulk Density

The following is a summary of methods for deriving canopy bulk density (CBD) (Mutlu, 2010; Agca et al., in review (a); Agca et al., in review (b)) defined as the density of available canopy fuels (Cruz et al., 2003; Scott and Reinhardt, 2001; Hall and Burke, 2006) using airborne lidar remote sensing technology (Task 4.6 – Crown bulk density or LMD:).

Height Bins Approach

The height bins approach (Popescu and Zhao, 2008) was used to generate a multiband dataset from the airborne lidar data to make use of the entire lidar point cloud. A total of eleven height bins were created and are illustrated in Figure 18. The first four height bins have vertical intervals of 0.5 m to afford a better characterization of vegetation that defines surface fuels. The fifth height bin contains points within 2.0 to 5.0 m. The next five height bins use 5 m intervals (5.0-10m, 10.0-15.0m, 15.0-20.0m, 20.0-25.0m, and 25.0-30.0m) and the final height bin contains all points above 30.0m. The three-dimensional height bins were converted to two-dimensional data by creating voxels, or volumetric pixels, (horizontal dimensions of 2.5 x 2.5 m and vertical dimension set by the height bin) and counting the occurrence of the number of lidar points within each voxel and normalizing by the total number of points in the vertical column. All of the height bins were used in this study to derive CBH and CBD canopy fuel parameters.

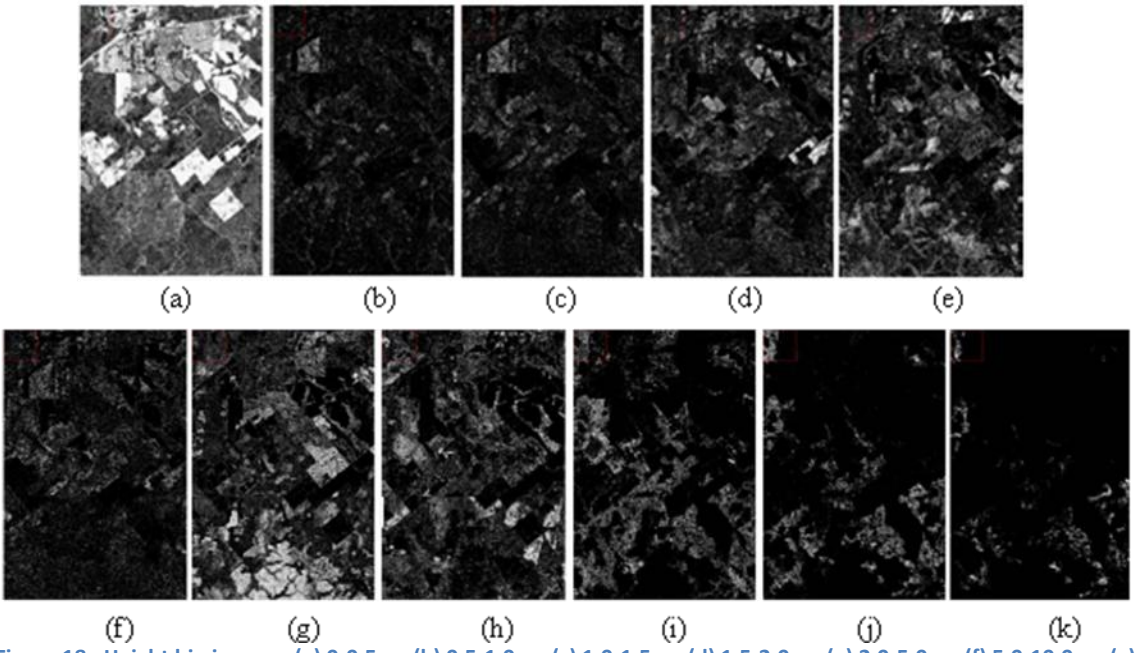


Figure 18. Height bin images: (a) 0-0.5 m, (b) 0.5-1.0 m, (c) 1.0-1.5 m, (d) 1.5-2.0 m, (e) 2.0-5.0 m, (f) 5.0-10.0 m, (g) 10.0-15.0 m, (h) 15.0-20.0 m, (i) 20.0-25.0 m, (j) 25.0-30.0 m, (k) >30.0 m.

Principal Component Analysis

A new dataset was created by stacking the four bands of the QuickBird image with all of the lidar height bins, creating a fifteen band image. To avoid bias in the stacked image, the four QuickBird bands were normalized and rescaled from their 2048 possible values to values between 0 and 1. Principal Component Analysis (PCA) was then performed on the fifteen band image to reduce the dimensionality of the dataset by transforming a set of correlated variables into a new set of uncorrelated variables (Jensen, 2005).

Eigenvalues, variance, and eigenvectors were extracted for each principal component (PC). Table 7 represents the eigenvalues, percentage of total variance, and cumulative variance explained by each PC. Each eigenvector represents a principle component. The first PC is defined by the eigenvector with the highest corresponding eigenvalue. The individual eigenvalues indicate the variance, the higher the value the more variance they have captured. The first five components, which account for approximately 92 % of the total variance, were used to derive CBD and CBH canopy fuel parameters. Figure 19 represents all the PCA components used in this study.

PCs	Eigenvalue	% of total variance	cumulative
PC1	0.124942	48.2299	0.482299
PC2	0.040637	15.6866	0.639165
PC3	0.032816	12.6676	0.765841
PC4	0.020207	7.8003	0.843844
PC5	0.01866	7.2031	0.915875

Table 7. Calculated percentage of total variance, eigenvalues, and cumulative variance explained by each principal component.

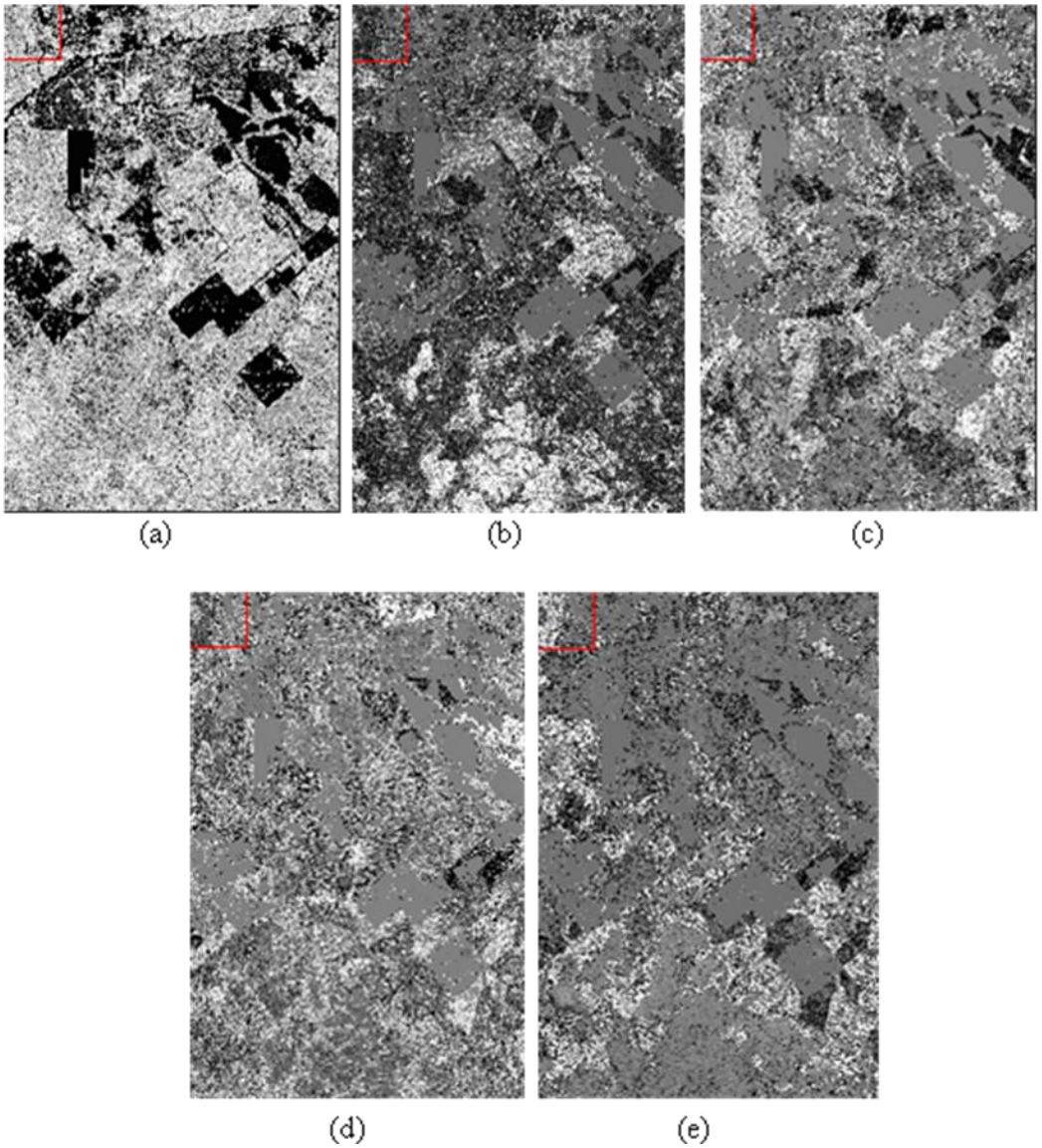


Figure 19. The first five PCs: (a) PC 1, (b) PC 2, (c) PC 3, (d) PC 4, and (e) PC 5.

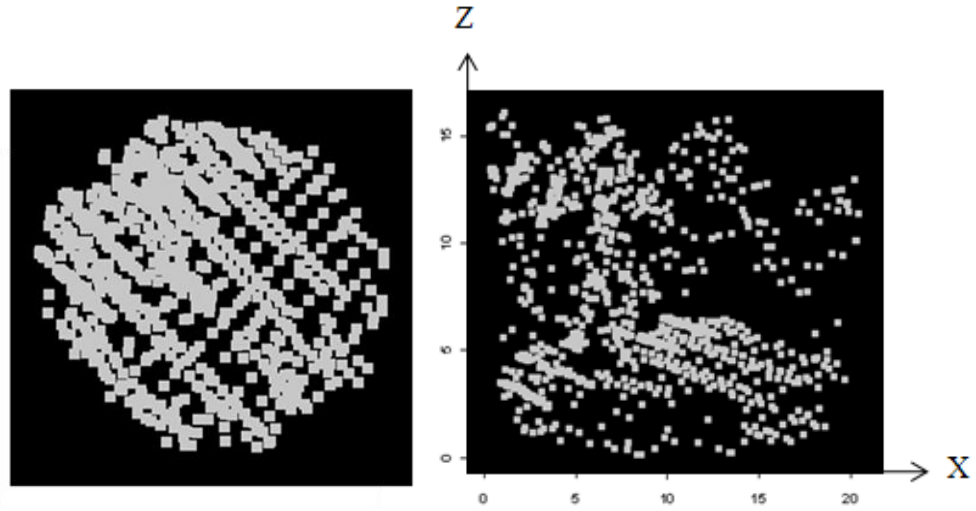
Deriving Canopy Fuel Parameters

Of the 50 field plots surveyed, 33 are 404.7 m² (1/10th acre) and 17 are 40.47 m² (1/100th acre). The smaller plot size was selected because the loblolly pine trees in unthinned plantations are uniform and the 40.47 m² (1/100th acre) size was enough to represent the stand. Four processing approaches were used to derive CBD and CBH canopy fuel parameters:

1. Two different datasets were used to derive these canopy fuel parameters from the entire airborne lidar point cloud at two different map spatial resolutions: (1) at 30 meter resolution, (2) at a grid cell size with an area equal to the actual plot size.
2. CBD and CBH were derived at plot level using an individual tree approach.
3. The upper lidar height bins corresponding to the canopy were used to derive these canopy fuel metrics.
4. PCA components (data fusion) were used.

Deriving CBD and CBH from lidar point cloud at different spatial resolutions

Similar to studies of Naesset and Bjerknes (2001), Erdody and Moskal (2010), and Andersen et al. (2005), eight lidar metrics were derived from the lidar point cloud including: 25th, 50th (median), 75th, and 90th of height percentiles of laser pulses, maximum height, mean height, coefficient of variation (cv), and canopy density (D) at plot size (20 m) and at 30 m horizontal grid size centered on plot center. The 30 m resolution was chosen to compensate for any GPS errors when locating ground plots. Quick Terrain Modeler (QTM) was used to derive the D metric at 20 m, while at 30 m resolution D was calculated as the number of all returns above 2.5 m divided by the total number of all returns within the 30 m pixel. In addition, logarithmic transformation was applied to our metrics. The log transformation is used for normalizing the data, easier data visualization, and correcting heteroscedasticity (Fox, 1997). This transformation also helps homogenize the variance over the sample data and equalizes the variance over the entire range of predicted *y* values (Sprugel, 1983). Including log-transformed metrics, a total of sixteen metrics were obtained for each resolution.



(a) (b)

Figure 20. Two snapshots of the Plot# Hunt_6 (r =11.35 m) lidar point cloud from the QTM software from above (a) and a side (b), respectively.

Deriving at plot level using an individual tree approach

Total tree height, crown ratio, foliage biomass, age, and crown length are the variables necessary to derive CBD using allometric equations, including: total above ground biomass (Equation 2), foliage biomass (Equation 3), crown volume (Equation 4), age (Equation 5), CBD (Equation 6), and Lorey's mean CBH (Equation 7).

$$Bm = exp(\beta_0 + \beta_1 \ln(dbh))$$

Equation 2. Jenkins et al. (2003) equation for total above ground biomass.

where Bm is total aboveground biomass (kg) for trees 2.5 dbh and larger, dbh is diameter at breast height (cm), Exp is exponential function, β_o is set to -2.5356, and β_1 is set to 2.4349.

$$FBratio = Exp(\beta_0 + \frac{\beta_1}{dbh})$$

Equation 3. Jenkins et al. (2003) equation for foliage biomass.

where FBratio is foliage biomass ratio to total aboveground biomass for trees 2.5 cm dbh and larger, β_o is equal to -2.9584, and β_1 is equal to 4.4766.

$$CRVOL = \pi FL[3b_1^2 - 5b_1b_2 + \frac{b_2^2}{3} - 4\ln(2) \times (b_1^2 - 2b_1b_2) + \frac{b_3^3}{3b_4}]$$

Equation 4. Baldwin and Peterson (1997) equation for crown volume.

The quantities b_1 through b_4 are defined by sub-models:

b1 = -4.5121 + 0.5176DBH + 4.3529R,

b2 = 4.4749 – 0.0175A – 0.4985DBH – 6.0414R,

b3 = 0.0168DBH + 0.0155FL, and

b4 = -0.0233dbh,

where A, R, and FL are the tree's area, crown ratio, and foliated crown length, respectively.

$$A = e^{\left(\frac{(TH + 0.453 \times SI)}{0.441 \times SI}\right)}$$

Equation 5. Stukey (2009) equation for tree age.

where TH is tree height and SI is site index.

$$CBD_{AL} = \frac{\Sigma FB}{\Sigma CV}$$

Equation 6. Equation for calculating canopy bulk density for the entire plot.

$$CBH_{LH} = \frac{\Sigma(BA \times CrBH)}{\Sigma BA}$$

Equation 7. Brack (1999) equation for Lorey's mean canopy base height.

where CBH_{LH} is canopy base height estimated using Lorey's mean height equation at plot level, CrBH is crown base height at individual tree level, and BA stands for basal area.

All of these variables were derived from airborne lidar data using the studies of Popescu and Zhao (2008) and Stukey (2009). Popescu and Zhao (2008) used TreeVaw (Tree Variable Window) developed by Popescu and Wynne (2004) to extract tree heights and crown diameters at individual tree level in our study area. Stukey (2009) used the results of Popescu and Zhao (2008) and Popescu (2007) to derive dbh, CBH, and age of each individual pine tree for the same study area. Thousands of trees for the whole study area were identified by the TreeVaw software. The locations of plots and all the trees captured by TreeVaw were displayed in ArcGIS 9.2. Then, trees within each plot boundaries were selected and a text file was created for further analysis. Of the 50 plots surveyed, only 41 were used because TreeVaw did not identify any trees for the other nine plots. In addition, foliage biomass and crown length of each tree were estimated using dbh and total tree height data obtained from Popescu and Zhao (2008) and Stukey (2009) to calculate CrBD at individual tree levels and CBD at plot levels.

Using the upper lidar height bins and PCA components to derive canopy fuel parameters

The seven upper lidar height bins (bins 5 through 11, Figure 18) were used to derive CBD and CBH canopy fuel parameters. The first four lidar height bins were not used since they represent the surface fuels (Mutlu et al., 2008). For each plot, thirty-five metrics were derived; including maximum, minimum, mean, standard deviation (st), and coefficient of variance (cv) of the digital numbers of each lidar bin at plot level. In addition, twenty-five metrics were derived from the PCA dataset including: maximum, minimum, mean, st, and cv of the digital numbers of each PCA component at plot level. ENVI 4.5 software was used to derive all the metrics. Then, all of the results were saved in a text file for regression analysis.

To estimate CBD and CBH canopy fuel parameters from airborne lidar data, multiple predictive models were developed in this study. In our regression models, we used the results of field measurements as inputs in the CBD and CBH equations (CBD-AL and CBH-LH) and CrownMass, a software package that estimates CBD and CBH, (CBD-CM and CBH-CM) as reference data. Four different metric sets were investigated to obtain the regression model of best fit for canopy fuels (summarized in Table 8): (1) sixteen lidar point cloud metrics at 30 m resolution, including the original eight lidar point cloud metrics and the same metrics transformed using a natural logarithmic transformation (Ln), named metrics-set-1, (2) a total of sixteen lidar point cloud metrics at actual plot size, including the original eight lidar point cloud metrics and the same metrics transformed using a natural logarithmic transformation (Ln), named metrics-set-2, (3) thirty-five metrics from the upper lidar height bins, named metrics-set-3, and (4) twenty-five metrics from the PCA dataset, named metrics-set-4.

Source	Definition
Metrics-set-1	sixteen lidar point cloud metrics at 30 m
Metrics-set-2	sixteen lidar point cloud metrics at actual plot size
Metrics-set-3	thirty-five lidar upper bins metrics
Metrics-set-4	twenty-five data fusion PCA metrics

Table 8. Definition of Metric sets used for this study

Table 9 represents all the independent variables used in the regression analysis. In this table, the subscript *i* for height bins metrics represents height bins 5 through 11 and the subscript *i* for data fusion metrics represents PCA bands 1 through 5, respectively. As mentioned before, log transformation was applied to all of lidar metrics. Log transformations

may introduce a systematic bias into the calculations; therefore, there is a need to calculate the correction factor to neutralize this bias. We calculated the correction factor using the following equation from Sprugel (1983):

$$Correction\ Factor = Exp(standard\ error\ of\ estimate)^2/2$$

Equation 8. Equation to neutralize bias introduced through log transformations.

The results of using field data in the allometric equations and CrownMass were used as ground data to see which method's result has a stronger relationship with the lidar derived metrics. Stepwise regression was performed to find the model of best fit for the data at $\alpha = 0.05$ for estimating CBD and CBH from airborne lidar. The selected models were chosen based on several criteria: a good balance between a high coefficient of determination (R^2) value, a low root mean square error (RMSE), no colinearity, and parsimony, which contains a limited number of independent variables. Variance inflation tests were conducted for each selected model which is important for finding colinearity between independent variables if it exists.

Metrics	Description
Lidar point cloud metrics	
h_mean	Mean height of point cloud
h_max	Max height of point cloud
h_25	25th percentile height of point cloud
h_50	50th percentile height of point cloud
h_75	75th percentile height of point cloud
h_90	90th percentile height of point cloud
D	Density
variance(s)	coefficient of variation of point cloud
CBH_{LH}	CBH obtained from allometric equations
CBH_{CM}	CBH obtained from CrownMass program
Height Bins metrics	
Bin_i_min	Lidar Height Bin_i: Minimum DN value of plot
Bin_i_max	Lidar Height Bin_i: Maximum DN value of plot
Bin_i_mean	Lidar Height Bin_i: Mean DN value of plot
Bin_i_st	Lidar Height Bin_i: Standard Deviation DN value of plot
Bin_i_cv	Lidar Height Bin_i: Variance DN value of plot
Data Fusion metrics	
PCA_i_min	PCA $Band_i$: Minimum DN value of plot
PCA_i_max	PCA $Band_i$: Maximum DN value of plot
PCA_i_mean	PCA $Band_i$: Mean DN value of plot
PCA_i_st	PCA $Band_i$: Standard Deviation DN value of plot
PCA_i_ cv	PCA $Band_i$: Variance DN value of plot

Table 9. All the metrics used in this study to derive CBD and CBH.

Results

Table 10 and Table 11 represent the regression models with a significance of 0.05 and their R^2, adjusted R^2, p-values, and RMSE values for both CBH and CBD, respectively. In these tables, $\hat{y}$ represents reference data for CBH and CBD and bias correction factors are added to the end of each predicted model.

As shown in Table 10, we obtained a total of eight models, two for each metrics set, of best fit for CBH. All the predicted regression models provided good R^2 values, ranging from 0.662 to 0.976 and adjusted R^2 values, ranging from 0.647 to 0.973. However, R^2 and adjusted R^2 values alone are not enough to select the best regression model; therefore, all the

selected models were plotted (Figure 21 (a) through Figure 21 (h)) to model the goodness-of-fit to choose the best model for predicting CBH. The models with the highest R^2 and adjusted R^2 values for all metric sets use CBH-LH as the reference data, with metrics-set-1 having the best result (Table 10).

Plot size 30 m	Model	R^2	Adj. R^2	P-value	RMSE
CBH_{LH}: Metrics-set-1	$\hat{y} = exp(1.326 + 1.001\,Ln(H_{mean}) - 0.036H_{50} - 1.21Ln(H_{90}) - 1.004D + (0.194)^2/2)$	0.976	0.973	<0.0001	0.195
CBH_{CM}: Metrics-set-1	$\hat{y} = exp(3.164 + 3.118Ln(cv) - 2.21D - 2.204Ln(H_{90}) + (0.268)^2/2)$	0.945	0.942	<0.0001	0.268
Plot size 20 m					
CBH_{LH}: Metrics-set-2	$\hat{y} = exp(-1.861 + 1.805LnH_{90} - 0.044H_{max} + (0.173)^2/2)$	0.949	0.947	<0.0001	0.173
CBH_{CM}: Metrics-set-2	$\hat{y} = exp(-1.358 + 1.21LnH_{90} + 0.086LnH_{25} + (0.234)^2/2)$	0.895	0.891	<0.0001	0.234
CBH from Metrics-set-3					
CBH_{LH}: Metrics-set-3	$\hat{y} = exp\,(-0.430 + 5.537Band7_{cv} + 6.31Band9_{mean} + (0.698)^2/2)$	0.708	0.696	<0.0001	0.698
CBH_{CM}: Metrics-set-3	$\hat{y} = exp(-0.431 + 4.747Band8_{cv} + 3.818Band6_{cv} + (0.715)^2/2))$	0.662	0.647	<0.0001	0.715
CBH from Metrics-set-4					
CBH_{LH}: Metrics-set-4	$\hat{y} = exp(0.079 + 4.634PCA1_{max} + 1.82PCA3_{mean} + 1.889PCA5_{mean} + (0.464)^2/2)$	0.874	0.866	<0.0001	0.464
CBH_{CM}: Metrics-set-4	$\hat{y} = exp\,(0.048 + 4.276PCA1_{max} + 2.36PCA3_{mean} + 1.658PCA5_{mean} + (0.450)^2/2)$	0.869	0.860	<0.0001	0.450

Table 10. Results of the CBH regression analysis.

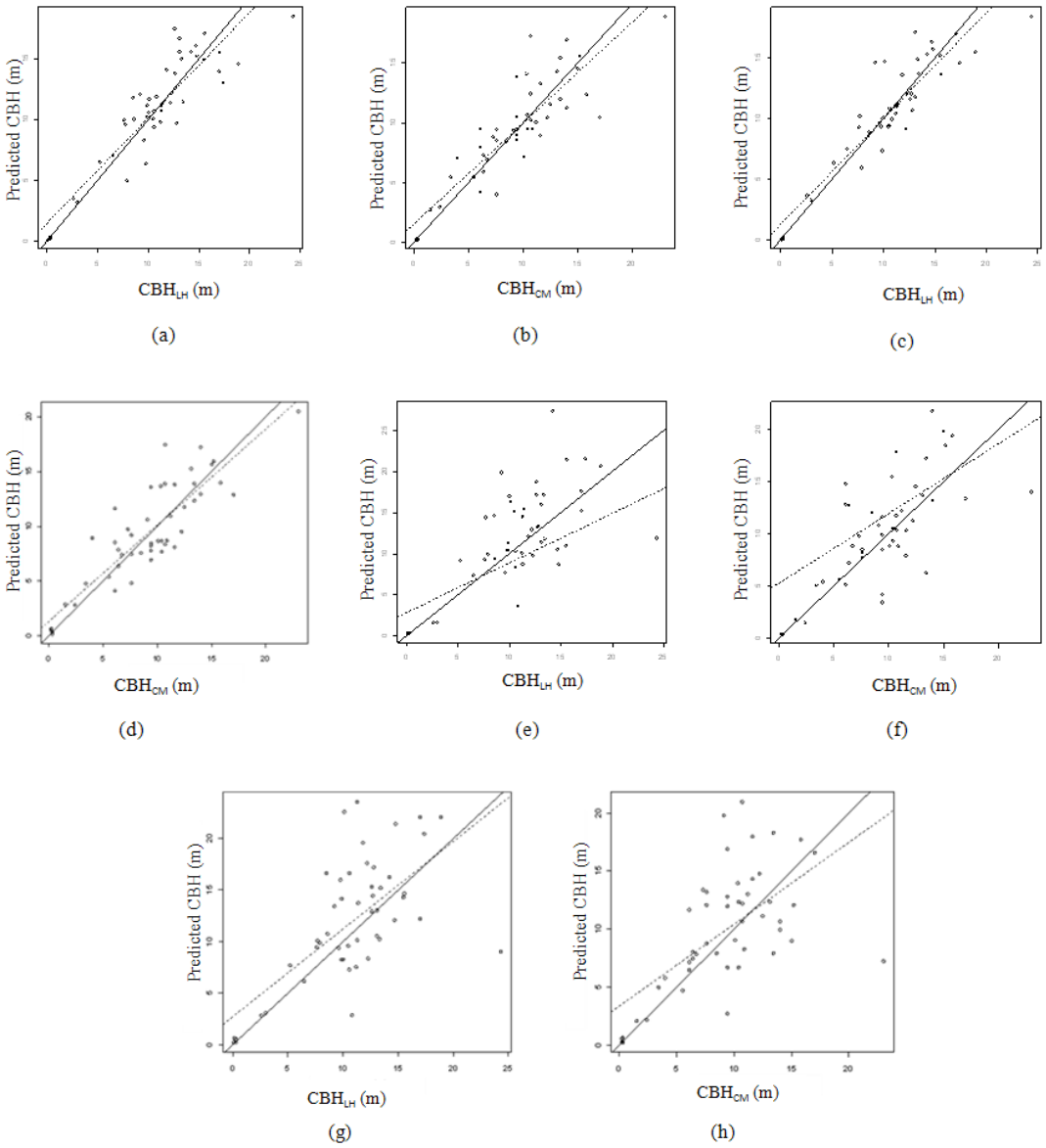


Figure 21. Scatter plots of predicted CBH vs. (a) CBH-equations at 30 m resolution; (b) CBH-CrownMass at 30 m resolution; (c) CBH-equations at actual plot size; (d) CBH-CrownMass at actual plot size; (e) CBH-equations from Metrics-set-3; (f) CBH-CrownMass from Metrics-set-3; (g) CBH-equations from Metrics-set-4; (h) CBH-CrownMass from Metrics-set-4. In all figures, the solid line represents X=Y and the dashed line represents the data fit.

As shown in Table 11, eight regression models~~,~~ (two for each metrics set)~~,~~ with a significance level of 0.05 were developed for CBD. These models were constructed using the same metrics sets and spatial resolutions as the regression models for CBH. The estimated CBH obtained previously was included as an additional variable. Bias correction factors are added to the end of each predicted model and ŷ represents ground validation data for CBD in Table 11. Again, the models using reference data from the allometric equations outperformed those using CrownMass. To illustrate the goodness-of-fit and select the best fitted model for predicting CBD, all regression models were plotted (Figure 22 (a) through Figure 22(h)). Overall, the best model to estimate CBD was obtained from Metrics-set-3 (CBD-AL), based on the scatter plots, pvalues, R^2, and adjusted R^2 values.

Plot size 30 m	Model	R^2	Adj. R^2	P-value	RMSE
CBD_{AL}: Metrics-set-1	$\hat{y} = \exp(-2.058 - 0.059H_{max} + 0.658\mathrm{Ln}(CBH) + (0.366)^2/2)$	0.660	0.645	<0.0001	0.366
CBD_{CM}: Metrics-set-1	$\hat{y} = \exp(-1.118 - 2.92D - 0.04H_{max} + (0.554)^2/2)$	0.473	0.450	<0.0001	0.554
Plot size 20 m					
CBD_{AL} : Metrics-set-2	$\hat{y} = \exp(-2.139 - 0.083H_{90} + 0.814\mathrm{Ln}(CBH) + (0.303)^2/2)$	0.689	0.674	<0.0001	0.303
CBD_{CM} : Metrics-set-2	$\hat{y} = \exp(-3.389 - 0.088H_{max} + 1.209\mathrm{Ln}(H_{mean}) + (0.386)^2/2)$	0.647	0.630	<0.0001	0.386
CBD from Metrics-set-2					
CBD_{AL} : Metrics-set-3	$\hat{y} = \exp(-2.558 + 0.585Band7_{max} + 0.414\mathrm{Ln}(CBH) - 1.37Band10_{cv} - 1.003Band8_{cv} + (0.329)^2/2)$	0.748	0.726	<0.0001	0.329
CBD_{CM} : Metrics-set-3	$\hat{y} = \exp(-3.733 + 1.492Band7_{max} + 1.593Band6_{mean} + (0.589)^2/2)$	0.403	0.378	<0.0001	0.589
CBD from Metrics-set-4					
CBD_{AL} : Metrics-set-4	$\hat{y} = \exp(-2.917 + 1.982PCA2_{cv} + 1.173PCA2_{min} + 2.179PCA4_{mean} + 0.285\mathrm{Ln}(CBH) + (0.359)^2/2)$	0.700	0.673	<0.0001	0.359
CBD_{CM} : Metrics-set-4	$\hat{y} = \exp(-2.986 + 1.145PCA1_{mean} - 5.414PCA5_{st} + 1.561PCA1_{max} + (0.516)^2/2)$	0.552	0.523	<0.0001	0.516

Table 11. Results of the CBD regression analysis.

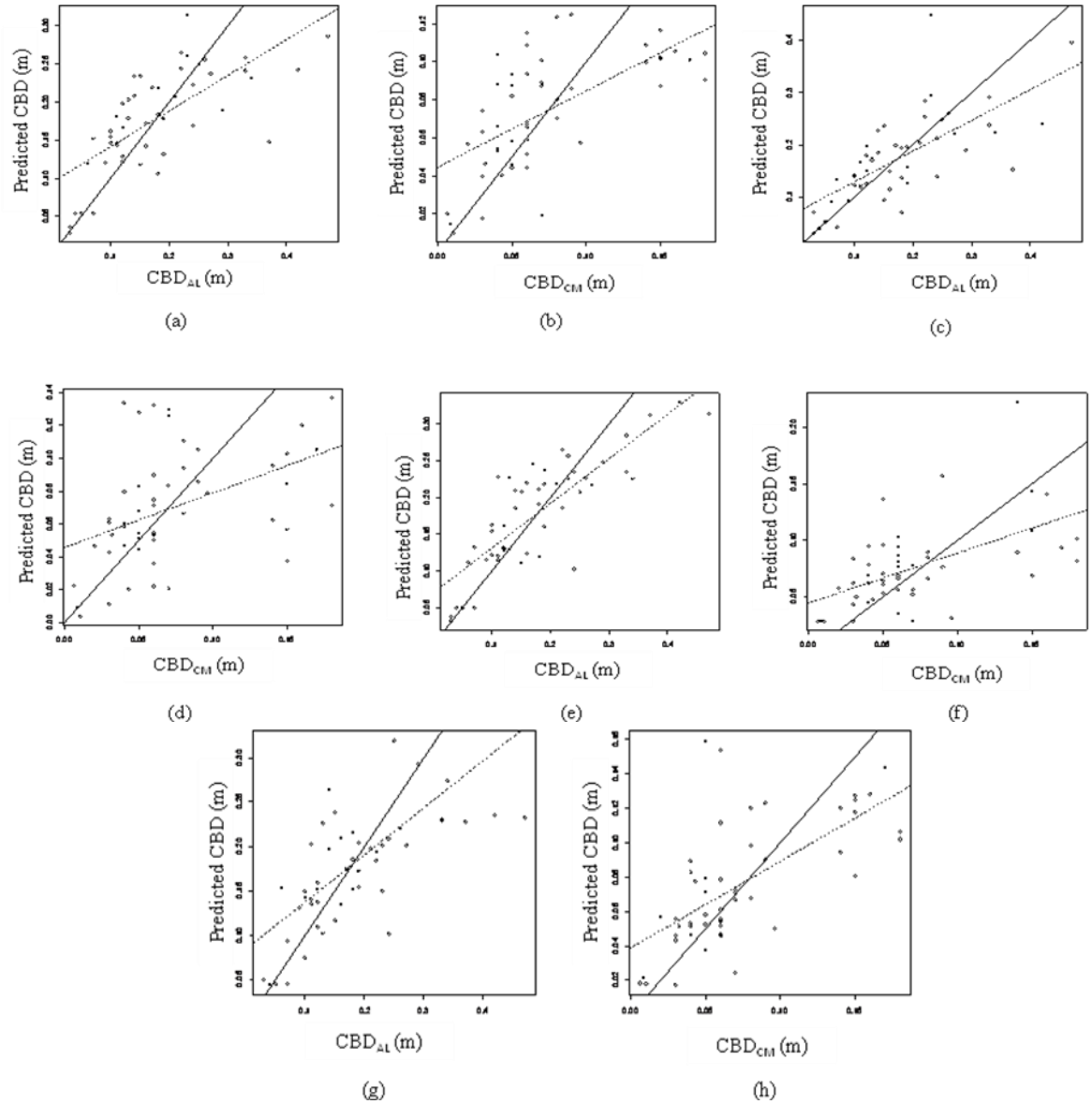


Figure 22. Scatter plots of predicted CBD vs. (a) CBD-equations at 30 m resolution; (b) CBD-CrownMass at 30 m resolution; (c) CBD-equations at actual plot level; (d) CBDCrownMass at actual plot level; (e) CBD-equations from Metrics-set-3; (f) CBD-CrownMass from Metrics-set-3; (g) CBD-equations from Metrics-set-4; (h) CBD-CrownMass from Metrics-set-4. In all figures, the solid line represents X=Y and the dashed line represents the data fit.

The results of predicted CBD and CBH from TreeVaw were plotted against CBD and CBH predicted through equations and CrownMass in Figure 23. Neither scatter plot for CBD shows a strong relationship; however, the CBD-TreeVaw and CBD-AL demonstrates a better relationship than the scatter plot of CBD-TreeVaw and CBD-CM. One possible explanation could be that TreeVaw captured fewer trees on each plot, affecting the result of CBD. There is a strong, positive relationship between CBH-TreeVaw and CBH-LH. Since there are uniform stands in the study area and we used weighted mean height approach to calculate CBH, the number of trees most likely does not affect the results.

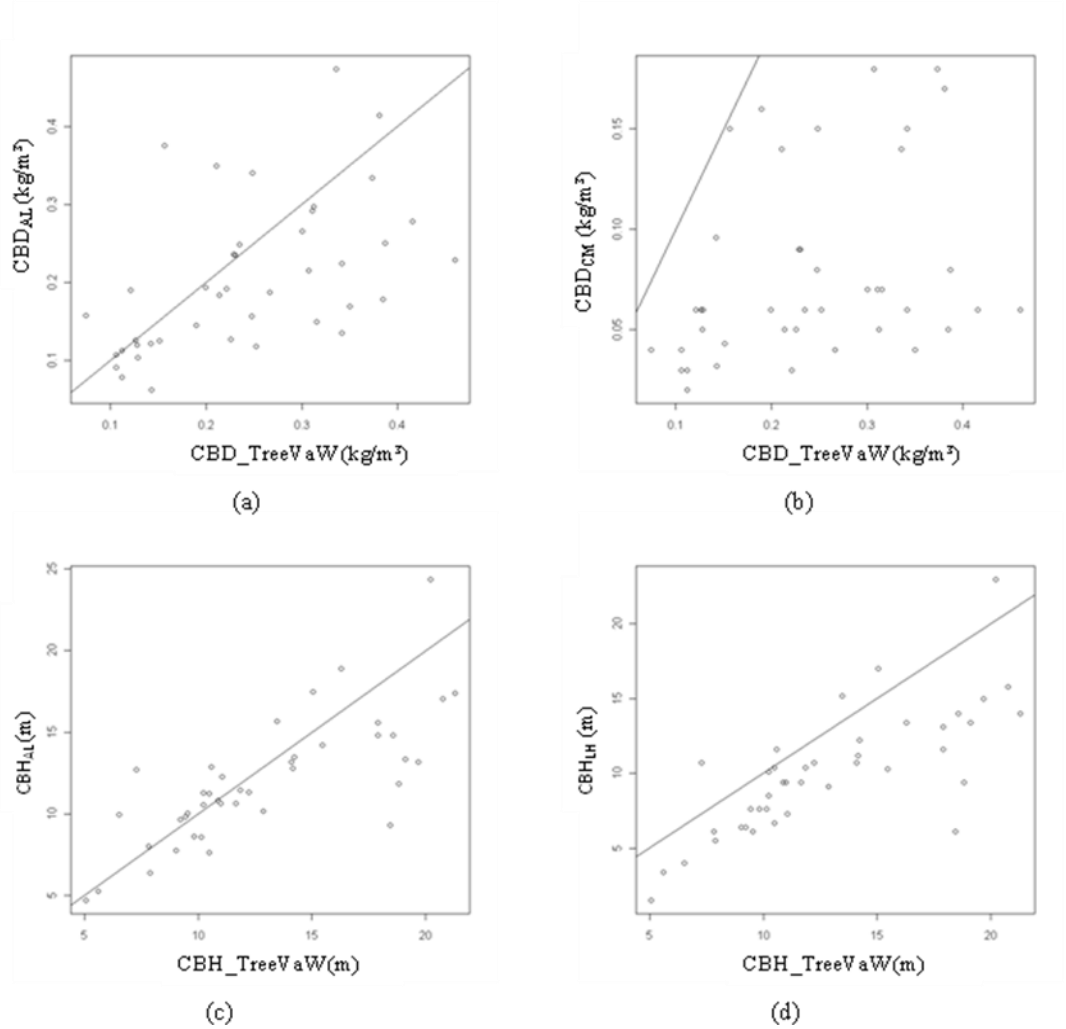
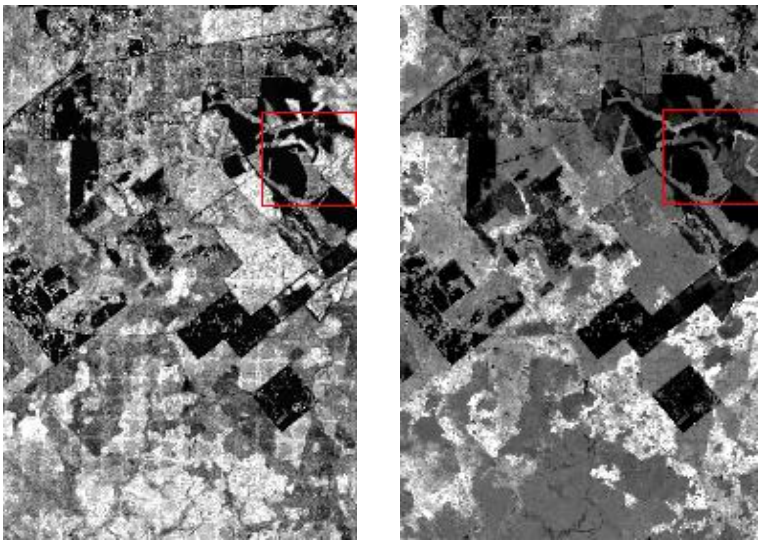


Figure 23. (a) CBD-equations versus predicted CBD from lidar, (b) CBD-CrownMass software versus predicted CBD from lidar, (c) CBH-equations versus predicted CBH from lidar, (d) CBH-CrownMass software versus predicted CBH from lidar.

After the regression models have been developed to establish a functional relationship between the airborne lidar data and the canopy fuels, CBD and CBH, the best models (metrics-set-3 CBD-AL and metrics-set-1 CBH-LH) were used to generate maps of CBD and CBH over the entire study area. The CBD (Figure 24Figure 24 (a)) and CBH (Figure 24 (b)) maps were generated in ENVI using the Band Math function based on the selected regression models.



(a) (b)

Figure 24. (a) The CBD map; (b) The CBH map of our study area.

Comparison of ground-based and airborne LAI estimations

Leaf Area Index (LAI) is the total one-sided area of green foliage per unit of ground surface and is an important variable for many process-based models, especially biogenic emissions models. Zhao and Popescu (2009) compared ground-based methods for estimating LAI using lidar and QuickBird satellite imagery for the Huntsville study area (Task 4.7 – Investigation of ground-based methodology for LAI and LMD assessment and comparison to airborne data:). The following is a summary of their findings. The full publication can be found at:

http://ssl.tamu.edu/media/14431/zhaopopescu%20rse%20lai%20august%202009.pdf

For the in-situ estimation of LAI, hemispherical photos were taken from the center of the plot and then processed through two different approaches. The first approach was having two separate analysts to use the commercial software package Hemiview (Delta-T Devices Ltd., UK, 1999) to evaluate the effect of subjectivity on hemispherical analysis caused by the manual and interactive selection of segmentation threshold values. The second approach used two different inversion algorithms based on the "gap-fraction" formula of the Beer's law in a canopy. Both algorithms assumed the leaf angle distribution is azimuthally-independent. The first algorithm uses an ellipsoidal distribution where the second algorithm does not restrict its functional form. Because there was a lack of ground-truth LAI data of higher accuracy, the most accurate approach could not be determined; therefore, the average of the four sets was used in the regression analysis relating in-situ LAI to the lidar-derived metrics.

Zhao and Popescu (2009) investigated laser penetration metrics, laser penetration metrics adjusted by look angle, height-related metrics, and foliage-density proxies. It was determined that a 25m plot diameter and 3.6m above ground for the canopy floor were the optimum model parameters. The laser penetration metrics were found to be the best of the lidar metrics, with an RMSE of 0.29 LAI when regressed with the in-situ LAI data. The lidar was then compared to the GLOBCARBON LAI product and found to be highly correlated with coefficients of 0.849 for pines and 0.854 for hardwoods. The initiative of GLOBCARBON (http://geofront.vgt.vito.be) is to deliver global parameters derived from multiple satellite image data for providing inputs to carbon cycle models (Plummer et al., 2006).

Zhao and Popescu (2009) shows lidar has promise in estimating LAI better than multispectral imagery. Misregistration between lidar-derived and GLOBCARBON LAI maps could severely bias pixel-by-pixel comparison, so registration is critical to direct comparison of the two LAI maps. Future studies should focus on collecting better in-situ LAI data to make a better comparison between lidar-derived LAI and actual LAI. Again, this is only an extremely brief highlight of the Zhao and Popescu (2009) article which should be read in its entirety to be fully understood.

New Data Acquisition

The lidar data used for the Huntsville study area is seven years old and needs to be updated. This project provided resources for acquisition of leaf-on lidar (Task 4.8 – 2009 LiDAR Acquisition:) and Hyperion hyperspectral data (Task 4.9 – Hyperion Acquisition:). The Huntsville study area contains mostly pines and does not provide a large population of broad-leafed trees. To overcome this, lidar and hyperspectral data were acquired for the Huntsville area and part of the Big Sandy Creek Unit of the Big Thicket National Forest east of Livingston, TX, about 65km east of the Huntsville area (Figure 25).

Terrapoint collected lidar data between October 21 and October 28, 2010 (leaf-on). The data was collected with a point spacing that provided a point density of 4 raw points per square meter with one return per pulse. Each pulse was capable of having 5 returns, leading to a possible 20 points per square meter maximum. Hyperion data for the Huntsville study area was unavailable so a scene including an area surrounding Liberty, TX was selected. Hyperion data was likewise unavailable for the Huntsville and Big Sandy Creek study areas. However, Airborne Real time Cueing Hyperspectral Enhanced Reconnaissance (ARCHER), which collects 52 bands of data between 500 and 1100nm at 1 meter resolution, was acquired from the Civil Air Patrol (CAP) to ensure the acquisition of hyperspectral data for these study areas.

The new lidar data collected from Terrapoint will be used to update the forest characteristics of the Huntsville study area as well as create a map of forest characteristics for the Big Sandy Creek Unit. The ARCHER data will be used to create a classification of individual tree species for the study areas. The combination of these two datasets will allow for better biogenic emissions estimations.

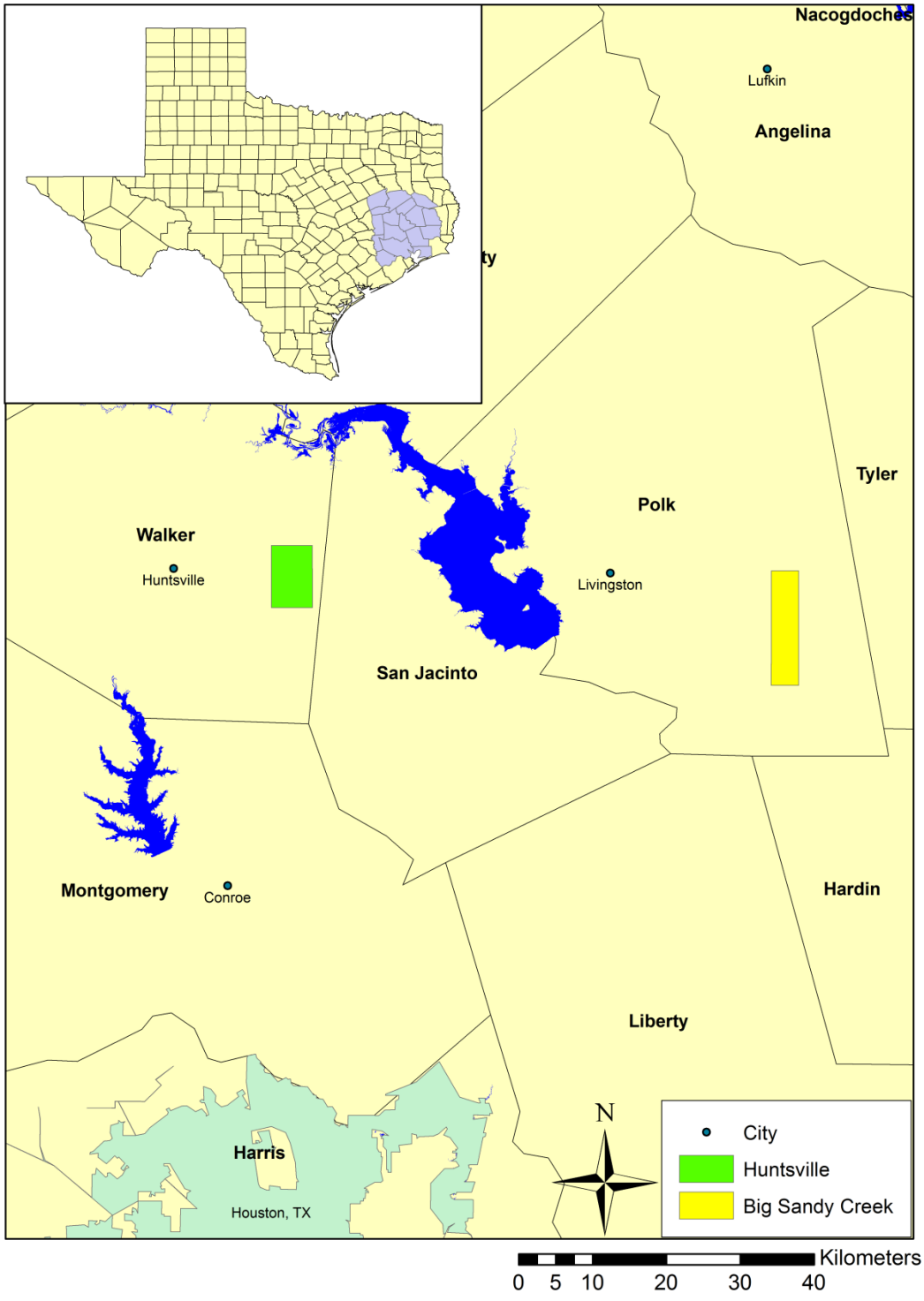


Figure 25. Map of study area locations for new lidar and ARCHER hyperspectral data acquisition.

Conclusions

The overall goal of all tasks was to provide more accurate and more informative datasets than currently available input to TCEQ's biogenic models. A LULC dataset was created by crosswalking existing datasets and classifying LandSat imagery, and lidar was used to map individual tree characteristics and estimate other variables for biogenic modeling.

The TCEQLC_2010 was produced by combining existing datasets to create a completely new dataset. The other benefit of using existing datasets, especially those produced by the USDA and Department of Interior, is that they already have accuracy assessments performed. The TCEQLC_2010 provides an extensive, detailed, and updatable LULC dataset that TCEQ can use for modeling air quality.

Lidar data were used to estimate individual tree characteristics of height, crown diameter, CBH, dbh, and location. These characteristics were then used to determine individual tree biomass, CBD, and surface roughness. Also, laser penetration metrics were used to determine leaf area index. This data and the products derived from it may prove to be highly useful in characterizing wooded and forested areas. Thus airborne lidar may prove to be an effective tool for land cover analysis useful in air quality modeling.

The project also delved into the new field of terrestrial lidar. While terrestrial lidar is limited in its coverage compared to airborne lidar, it is superior for directly measureing dbh, the most commonly used characteristic to estimate biomass. Terrestrial lidar also shows promise as a sampling tool to supplement airborne lidar data. Terrestrial lidar can save money by capturing the structure of field plots and reducing or eliminating the need for return visits by other field crews to verify accuracy.

This project benefited from the acquisition of lidar data in 2004 by the PI for a different project. The new acquisition of lidar and hyperspectral data will be used in future projects to analyze individual tree characteristics by tree species. Not only was data collected for the Huntsville area, but also for the Big Sandy Creek Unit of the Big Thicket National Preserve, which is more data than was originally projected for acquisition. The Big Sandy Creek is comprised of mostly broad-leafed species, which are of greater interest to air quality modelers. Broad-leafed species produce greater highly reactive biogenic emissions than needle-leafed species, and so their spatial distribution and identification is important. High resolution, hyperspectral data will make classifying individual trees at the species level possible, and lidar data can then be used to estimate the other physical variables valuable to air quality modeling.

References

Agca, M., Popescu, S.C., and Harper, C.W. (in review(a)). Deriving Forest Canopy Fuel Parameters for Loblolly Pine Trees in Eastern Texas. Submitted to Canadian Journal of Forest Research.

Agca, M., Popescu, S.C., Kaiguang, Z., and Eriksson, M (in review(b)). Estimating Canopy Bulk Density and Canopy Base Height Fuel Parameters using Airborne Lidar Data. Submitted to Remote Sensing of Environment.

Andersen, H. E., McGaughey, R. J., & Reutebuch, S. E. (2005). Estimating forest canopy fuel parameters using lidar data. *Remote Sensing of Environment, 94*, 441-449.

Baldwin, V. C. & Peterson, K. D. (1997). Predicting the crown shape of loblolly pine trees. *Canadian Journal Forest Research, 27*, 102-107.

Brack, C. (1999). Lorey's mean height, forest measuring and modeling. The Australian National University. http://srs-associated.anu.edu.au/mensuration/s_height.htm (Accessed on December 7, 2009).

Cruz, M. G., Alexander, M. E., & Wakimoto, R. H. (2003). Assessing canopy fuel stratum characteristics in crown fire prone fuel types of western North America. *International Journal of Wildland Fire, 12*, 39-50.

Erdody T., & Moskal, L. M. (2010). Fusion of lidar and imagery for estimating forest canopy fuels, *Remote Sensing of Environment*, *114*, 725-737.

Farr, T. G., et al. (2007), The Shuttle Radar Topography Mission, Rev. Geophys., 45, RG2004, doi:10.1029/2005RG000183.

Fox, J. (1997). *Applied regression analysis, linear models, and related methods.* Thousand Oaks, CA: Sage Press.

Hall, S. A., & Burke, I C. (2006). Considerations for characterizing fuels as inputs for fire behavior models. *Forest Ecology and Management, 227*, 102-114.

Jenkins, J. C., Chojnacky, D. C., Heath, L. S., & Birdsey, R. A. (2003). National-scale biomass estimators for United States tree species. *Forest Science*, *49*, 12-35.

Jensen, J. R. (2005). *Introductory digital image processing: A remote sensing perspective*. Upper Saddle River, NJ: Prentice–Hall Press.

Mutlu, M. (2010). Estimating canopy fuel parameters with in-situ and remote sensing data. Ph.D. Thesis. Texas A&M University, College Station, TX.

Mutlu, M., Popescu, S. C., Stripling, C., & Spencer, T. (2008). Mapping surface fuel models using lidar and multispectral data fusion for fire behavior. *Remote Sensing of Environment, 112*, 274-285.

Naesset, E., & Bjerknes, K. O. (2001). Estimating tree heights and number of stems in young forest stands using airborne laser scanner data. Remote Sensing of Environment, 78, 328 – 340.

Plummer, S., Arino, A., Simon, M., & Steffen, W. (2006). Establishing an Earth Observation Product service for the terrestrial carbon community: The Globcarbon initiative. Mitigation and Adaptation Strategies for Global Change, 11, 97−111.

Popescu, S. C., & Wynne, R. H. (2004). Seeing the trees in the forest: Using LIDAR and multispectral data fusion with local filtering and variable window size for estimating tree height. *Photogrammetric Engineering & Remote Sensing, 70*, 589–604.

Popescu, S. C. (2007). Estimating biomass of individual pine trees using airborne lidar. *Biomass and Bioenergy, 31,* 646-655.

Popescu, S. C. & Zhao, K. (2008). A voxel-based lidar method for estimating crown base height for deciduous and pine trees. *Remote Sensing of Environment, 112*, 767-781.

Scott, J. H., & Reinhardt, E. D. (2001). *Assessing crown fire potential by linking models of surface and crown fire behavior. General Technical Report RMRS-RP-29.* Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station 59 pp.

Sprugel, D. G. (1983). Correcting for bias in log-transformed allometric equations. *Journal of Ecology, 64*, 209-210.

Stukey, J. (2009). Deriving a Framework for Estimating Individual Tree Measurements with lidar for use in the TAMBEETLE Southern Pine Beetle Infestation Growth Model. M.S. thesis, Texas A&M University, College Station, TX.

Texas Geographic Information Council, 1999. Texas Land Classification System. Recommendations for New Land Use Land Cover Datasets for Texas.

Zhao, K., S. C. Popescu, & R.F. Nelson, 2009. Lidar remote sensing of forest biomass: A scale-invariant estimation approach using airborne lasers. Remote Sensing of Environment 113(1): 182-196.

Zhao, K. and S. C. Popescu, 2009. Lidar-based mapping of leaf area index and its use for validating GLOBCARBON satellite LAI product in a temperate forest of the southern USA. Remote Sensing of Environment 113 (8): 1628–1645.

APPENDIX A

Task 3 – Identification of LULC classes that have analogues with the NLCD vs. classes without direct correlation to NLCD:

Grantee shall conduct any and all analysis of the NLCD dataset, the TX LULC data, and pertinent ancillary data, to identify all land use/land cover layers that may be "crosswalked" with the NLCD as well as those layers that do not have a direct correlation to the NLCD data. A contact within the TCEQ modeling staff that is knowledgeable of the biogenic modeling shall be available for consultation during execution of this task. The grantee shall produce a technical memo, which shall detail which layers will and will not be crosswalked, how they were chosen and shall also include tables for cross-referencing layers between the NLCD and the TX LULC.

Task 3.1 – Identification and integration of ancillary GIS data:

Grantee shall identify relevant GIS data sets from state and federal agencies, academic sources, etc., and shall combine these data sets of tree species habitat and state-specific geospatial information into an integral whole, within the extent of the 12km CAMx grid. The grantee shall provide to the TCEQ project manager – in ArcGIS compatible format – all ancillary GIS data to be used in this task, unless such data is barred from further distributions by its originator

Task 3.2 – Field Work:

Grantee shall conduct field work at the Huntsville Texas A&M study area, to improve the accuracy of vegetation composition within the various land cover classes, and to validate ancillary datasets used in the crosswalk. Data collected may include diameter at breast height, tree locations, tree species, etc. Grantee may collect ground-based LiDAR data from various arboreal land cover classes and locations while performing this task. This will provide an opportunity to collect data for tasks 4.0-4.7. Grantee shall include an addendum to the May monthly report summarizing locations visited and type of data collected.

Task 3.3 – Supplemental Imagery Classification:

LANDSAT and other remote sensing imagery shall be obtained and classified on an as-needed basis for the purpose of distinguishing classes where the vegetation composition in the NLCD does not have an analogous class in the TX LULC, or where ascertaining subclasses. Grantee shall include an addendum to the June monthly report summarizing from which areas imagery was collected and from which time period the imagery was taken, and the source of each image.

Task 3.4 – Conversion Routines for converting NLCD to TX LULC

Grantee shall develop operable GIS routines that will be used for converting NLCD classes to TX LULC classes. Routines shall be written for either ENVI or ArcGIS, and shall be delivered to the TCEQ project manager along with any ancillary data necessary for the conversion of NLCD data to TX LULC data. Code must be demonstrated by the grantee to be operable on TCEQ hardware by the project manager.

Task 3.5 – Final Classification and Shapefiles:

Grantee shall deliver to the TCEQ project manager a final classified land use/land cover raster dataset resulting from the NLCD to TX LULC conversion, for the TCEQ 12km CAMx domain. This final classified dataset shall be accompanied by an

accuracy assessment. The grantee shall also deliver the final LULC classification as shapefiles, subset by County or Parish.

Task 4.0 – Canopy Height Model

Grantee shall generate the canopy height model (elevation data for the canopy height) for the Huntsville study area as derived from the 2004 LiDAR data. Grantee shall post the canopy height model for the study area for download via ftp by the TCEQ project manager, or else deliver it via external hard drive.

Task 4.1 – Above-ground Biomass Map:

Grantee shall generate an above-ground biomass map of the Huntsville study area, and shall deliver this biomass map to the TCEQ project manager via ftp or external hard drive.

Task 4.2 – Surface Roughness:

Using the 2004 LiDAR data, the grantee shall derive surface roughness for the Huntsville study area, and shall deliver the surface roughness measurements to the TCEQ project manager.

Task 4.3 – Ground-based LiDAR field measurements:

Grantee shall conduct field measurements of canopy base height, dbh, and canopy structure metrics with a ground-based LiDAR system. Grantee shall deliver a supplemental report to the April monthly report summarizing the measurements taken, and how they will be used in the scope of the project.

Task 4.4 – Maps of individual tree crown dimensions:

Grantee shall generate maps of the Huntsville study site indicating individual tree crowns and their dimensions, and shall provide some indication of the accuracy of the measurements. Grantee shall deliver the tree crown map to the TCEQ project manager via ftp or portable hard drive.

Task 4.5 – Comparison of ground and airborne LiDAR measurements compared

Grantee shall compare ground LiDAR and airborne LiDAR measurements and shall determine the best uses for both types of measurements. Grantee shall provide an in-depth analysis of the findings in a supplemental report to the June monthly report.

Task 4.6 – Crown bulk density or LMD:

Grantee shall derive crown bulk density or LMD measurements from airborne data for the Huntsville study site. These measurements shall be delivered to the TCEQ via ftp or external hard drive.

Task 4.7 – Investigation of ground-based methodology for LAI and LMD assessment and comparison to airborne data:

Grantee shall investigate the ground-based methodology for LAI and LMD assessment, and shall make comparisons to airborne data for the Huntsville study site. The focus of the comparison shall be on how closely airborne data matches

the ground-based methodology used for calculating LAI and LMD. Grantee shall include detailed results of the investigation in the Final Report to the TCEQ project manager.

Task 4.8 – 2009 LiDAR Acquisition:

Grantee shall coordinate with the TCEQ project manager to select a study area to be flown with lidar, and shall have the selected area flown. This new 2009 lidar data shall be delivered via ftp or external hard drive to the TCEQ project manager.

Task 4.9 – Hyperion Acquisition:

Grantee shall acquire Hyperion imagery in the study area covering the new 2009 lidar data, as well as the Huntsville study area. Hyperion is a hyperspectral sensor onboard the EO-1 NASA satellite, and its imagery will be useful for distinguishing arboreal vegetation spectrally. The imagery shall be acquired during the leaf-on period, and shall have <10% cloud cover, and preferable cloud-free. The Hyperion data shall be delivered via ftp or external hard drive to the TCEQ project manager.

APPENDIX B

```
# County Subset
# Jared Stukey
# Texas A&M University - Spatial Sciences Lab
# Subsets raster dataset by each record in shapefile and converts resulting
# raster to a shapefile

#####################################################
# plan is to select each record of the shapefile    #
# and convert the resulting raster to shapefile     #
# 1.  select by attributes using counter as FID     #
# 2.  get value of country, state, and county       #
# 3.  set workspace to country/ for mexico and      #
#     country/state/ for usa                        #
# 4.  replace spaces from county with underscores   #
# 5.  convert resulting raster to shapefile with    #
#     same naming convention                        #
#####################################################
# import operating system
import os

# import shutil to clear temp folder
import shutil

# import and create geoprocessor
import arcgisscripting
gp = arcgisscripting.create(9.3)

# import dbf reader
from dbfpy import dbf

# set workspace and dbf file
workspace = 'C:\\WORKSPACE\\TCEQ\\final'
dbffile = workspace + '\\TXLCS_counties.dbf'
db = dbf.Dbf(dbffile)

# add data management, conversion, and spatial analyst toolboxes
gp.AddToolbox("C:/Program Files (x86)/ArcGIS/ArcToolbox/Toolboxes/Spatial Analyst Tools.tbx")
gp.AddToolbox("C:/Program Files (x86)/ArcGIS/ArcToolbox/Toolboxes/Data Management Tools.tbx")
gp.AddToolbox("C:/Program Files (x86)/ArcGIS/ArcToolbox/Toolboxes/Conversion Tools.tbx")

# set local variable
gp.MakeFeatureLayer('C:\\WORKSPACE\\TCEQ\\final\\TXLCS_counties.shp', 'lyr')

# set environment for spatial analyst
if os.path.exists('C:\\temp\\scratch'):
   print "exists"
   gp.ScratchWorkspace = 'C:\\temp\\scratch'
else:
   os.makedirs('C:\\temp\\scratch')
   gp.ScratchWorkspace = 'C:\\temp\\scratch'
inraster = workspace + '\\TCEQLC_2010.tif'
```

```
# start counter for FID
i = 0
query = "FID = " + str(i)

# error trap
try:

   # iterate through records
   for r in db:
      # select layer by FID
      gp.SelectLayerByAttribute_management('lyr', "NEW_SELECTION", query)

      # set variables for directory and file name
      country = r['Country']
      state = r['STATE']
      state_nospace = state.replace(' ', '')
      county = r['COUNTY']
      county_nospace = county.replace(' ', '')[:-6]
      print country, state, county, state_nospace, county_nospace

      # set file structure
      if country == 'Mexico':
         if os.path.exists(workspace + '\\' + country):
            gp.workspace = workspace + '\\' + country
         else:
            os.makedirs(workspace + '\\' + country)
            gp.workspace = workspace + '\\' + country
      else:
          if os.path.exists(workspace + '\\' + country):
            if os.path.exists(workspace + '\\' + country + '\\' + state_nospace):
               gp.workspace = workspace + '\\' + country + '\\' + state_nospace
            else:
               os.makedirs(workspace + '\\' + country + '\\' + state_nospace)
               gp.workspace = workspace + '\\' + country + '\\' + state_nospace
          else:
            os.makedirs(workspace + '\\' + country + '\\' + state_nospace)
            gp.workspace = workspace + '\\' + country + '\\' + state_nospace


      # copy selected features
      if country == 'Mexico':
         gp.CopyFeatures('lyr', gp.ScratchWorkspace + '\\' + state_nospace)
      else:
         gp.CopyFeatures('lyr', gp.ScratchWorkspace + '\\' + state_nospace)

      # set variables for extract by mask
      mexoutraster = gp.ScratchWorkspace + '\\a' + str(i)
      usaoutraster = gp.ScratchWorkspace + '\\a' + str(i)
      print mexoutraster, usaoutraster
      inmask = gp.ScratchWorkspace + '\\' + state_nospace + '.shp'
      print inmask

      # check out spatial analyst extension license
```

```
    gp.CheckOutExtension("Spatial")

    # set analysis extent
    gp.Extent = inmask
    # extract by mask and convert to polygon
    if country == 'Mexico':
       shpworkspace = gp.workspace + '\\' + state_nospace + '.shp'
       gp.ExtractByMask_sa(inraster, inmask, mexoutraster)
       gp.RasterToPolygon_conversion(mexoutraster, shpworkspace, "NO_SIMPLIFY", "VALUE")
    else:
       shpworkspace = gp.workspace + '\\' + county_nospace + '.shp'
       print shpworkspace
       gp.ExtractByMask_sa(inraster, inmask, usaoutraster)
       gp.RasterToPolygon_conversion(usaoutraster, shpworkspace, "NO_SIMPLIFY", "VALUE")

    # clear scratch workspace
    shutil.rmtree(gp.ScratchWorkspace)

    #recreate scratch workspace
    if os.path.exists('C:\\temp\\scratch'):
       print "exists"
       gp.ScratchWorkspace = 'C:\\temp\\scratch'
    else:
       os.makedirs('C:\\temp\\scratch')
       gp.ScratchWorkspace = 'C:\\temp\\scratch'

       #update counter
       i += 1
       print i
  # If an error occurred print the message to the screen
except:
  print gp.getmessages()
#   print "something is wrong”
```